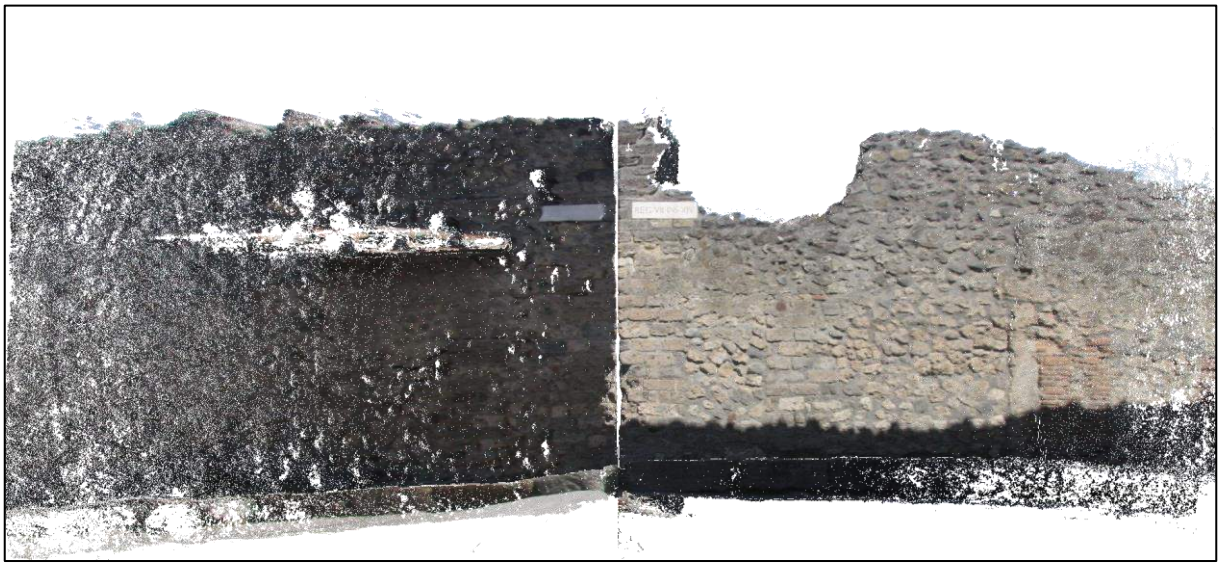
Ph.D. Program in Civil, Chemical and Environmental Engineering

Curriculum in Fluid Dynamics and Environmental Engineering

Department of Civil, Chemical and Environmental Engineering

Polytechnic School, University of Genoa, Italy



Photogrammetric suite to manage the survey workflow in challenging environments and conditions

Sara Gagliolo

PHOTOGRAMMETRIC SUITE TO MANAGE THE SURVEY WORKFLOW IN CHALLENGING ENVIRONMENTS AND CONDITIONS

BY

SARA GAGLIOLO

Dissertation discussed in partial fulfilment of
the requirements for the Degree of

DOCTOR OF PHILOSOPHY

Civil, Chemical and Environmental Engineering
curriculum in Fluid Dynamics and Environmental Engineering,
Department of Civil, Chemical and Environmental Engineering, University of Genoa, Italy

September, 2021

Adviser:

Prof. Domenico Sguerso – Department of Civil, Chemical and Environmental Engineering, University of Genoa, Italy

External Reviewers:

Prof. Eija Honkavaara – Department of Remote Sensing and Photogrammetry, National Land Survey of Finland, Finland

Prof. María Clara Patricia de Lacy Pérez de los Cobos – Department of Cartographic, Geodetic and Photogrammetric Engineering, University of Jaen, Spain

Examination Committee:

Prof. Matti Vaaja – Department of Built Environment, Aalto University, Finland

Prof. Vittorio Casella – Department of Civil Engineering and Architecture, University of Pavia, Italy

Prof. Rodolfo Repetto – Department of Civil, Chemical and Environmental Engineering, University of Genoa, Italy

Ph.D. program in Civil, Chemical and Environmental Engineering

Curriculum in Fluid Dynamics and Environmental Engineering

Cycle XXXIII

Cover-designers: Giulia Curletto (XXIX cycle) and Margherita Pongiglione (XXVIII cycle)

ABSTRACT

The present work is intended in providing new and innovative instruments to support the photogrammetric survey workflow during all its phases. A suite of tools has been conceived in order to manage the planning, the acquisition, the post-processing and the restitution steps, with particular attention to the rigorousness of the approach and to the final precision.

The main focus of the research has been the implementation of the tool MAGO, standing for Adaptive Mesh for Orthophoto Generation. Its novelty consists in the possibility to automatically reconstruct “unrolled” orthophotos of adjacent façades of a building using the point cloud, instead of the mesh, as input source for the orthophoto reconstruction.

The second tool has been conceived as a photogrammetric procedure based on Bundle Block Adjustment. The same issue is analysed from two mirrored perspectives: on the one hand, the use of moving cameras in a static scenario in order to manage real-time indoor navigation; on the other hand, the use of static cameras in a moving scenario in order to achieve the simultaneously reconstruction of the 3D model of the changing object.

A third tool named U.Ph.O., standing for Unmanned Photogrammetric Office, has been integrated with a new module. The general aim is on the one hand to plan the photogrammetric survey considering the expected precision, computed on the basis of a network simulation, and on the other hand to check if the achieved survey has been collected compatibly with the planned conditions. The provided integration concerns the treatment of surfaces with a generic orientation further than the ones with a planimetric development.

After a brief introduction, a general description about the photogrammetric principles is given in the first chapter of the dissertation; a chapter follows about the parallelism between Photogrammetry and Computer Vision and the contribution of this last in the development of the described tools. The third chapter specifically regards, indeed, the implemented software

and tools, while the fourth contains the training test and the validation. Finally, conclusions and future perspectives are reported.

INDEX

LIST OF ACRONYMS

AGL: Above Ground Level

AI: Artificial Intelligence

BBA: Bundle Block Adjustment

CCD: Charge-Coupled Device

CMOS: Complementary Metal Oxide Semiconductor

CNN: Convolutional Neural Networks

CRF: Conditional Random Fields

CV: Computer Vision

DLT: Direct Linear Transformation

DOF: Degrees Of Freedom

DSM: Digital Surface Model

EKF: Extended Kalman Filter

EO: External Orientation

FOD: First Order Design

FPS: Frame (Rate) Per Second

GCP: Ground Control Points

GSD: Ground Sample Distance

GUI: Graphical User Interface

HSV: Hue Saturation Value

IMU: Inertial Measurement Unit

IO: Internal Orientation

LIDAR: Light Detection and Ranging

LO: Longitudinal Overlapping

LS: Laser Scanner

M3C2: Multiscale Model to Model Cloud Comparison

MAGO: Adaptive Mesh for Orthophoto Generation

ML: Machine Learning

MRF: Markov Random Fields

NAN: Not A Number

NCC: Normalized Cross-Correlation

RANSAC: RANdom SAmpling Consensus

RF: Reference Frame

RFID: Radio-Frequency Identification

RGB: Red Green Blue

RMSE: Root Mean Square Error

RS: Reference System

SAD: Sum of Absolute Difference

SFM: Structure From Motion

SIFT: Scale Invariant Feature Transform

SLAM: Simultaneous Localization And Mapping

SOD: Second Order design

SONAR: Sound Navigation and Ranging

SOR: Statistical Outlier Removal

SPP: Spatial Pyramid Pooling

SSD: Sum of Squared Differences

SVD: Singular Value Decomposition

TLS: Terrestrial Laser Scanner

TO: Transversal Overlapping

TOD: Third Order Design

TOF: Time Of Flight

TS: Total Station

UAV: Unmanned Aerial Vehicle

U.Ph.O.: Unmanned Photogrammetric Office

UWB: Ultra-Wide Band

VLC: Visible Light Communications

VLF: Very-Low Frequency

ZOD: Zero-Order Design

LIST OF FIGURES

APPENDIX

LIST OF TABLES

APPENDIX

INTRODUCTION

In the last decades, the increasing spread of UAVs (Unmanned Aerial Vehicles) in civilian applications enhanced, among others, the potentialities of Photogrammetry in obtaining good results in terms of speed, cheapness, precision and accuracy, assuring at the same time the safety of the operators even in critical situations (e.g. natural disasters). This contribution made the photogrammetric survey technique suitable for a wider range of scenarios, previously restricted or not optimal in case of terrestrial or aerial Photogrammetry.

The growing interest of the scientific community about Photogrammetry, thanks to its aforementioned peculiarities and feasibility in combination with UAVs, led to the parallel need of software solutions, in order to improve the management of the whole workflow, including the planning, the acquisition, the post-processing and the restitution steps. Against this background, the aim of the research work during the PhD triennium has been to optimize the operative phases in terms of quality, precision, time, and computational effort, by means of dedicated tools focused on giving to a moderately skilled user the instruments to custom the processing, according to his final objective.

The best-known commercial and open source software solutions are based on both Photogrammetry and Computer Vision (CV) principles and algorithms. In this regard, the contribution of CV and, in general, of image processing has been very fundamental even in the implementation approach of the developed tools, in order to get a higher level of automation without the loss of the statistical meaning. A joint view of Photogrammetry and CV has been embraced, since these two fields recently converged, in order to find generic solutions for orientation and reconstruction problems.

The outcome of the present work consists in a suite of tools. The first, MAGO (Adaptive Mesh for Orthophoto Generation), is implemented in C++ environment; its novelty is the possibility to automatically reconstruct joint orthophotos of adjacent façades of a building by using the point cloud, instead of the polygonal mesh, as input source for the orthophoto reconstruction.

A second tool, which development has begun in C++ environment but it is still ongoing because of the hindrance of Covid-19 pandemic outbreak, has been conceived to allow the management of two mirrored situations, relying on a photogrammetric procedure using Bundle Block Adjustment. The former scenario considers the use of moving cameras in a static scene in order to let real-time navigation and positioning in such environments not reachable by the GNSS signal; the latter permits the simultaneous restitution of the 3D model of a moving object surveyed by means of static cameras.

The third tool, named U.Ph.O. (Unmanned Photogrammetric Office), has been mostly implemented by Eng. Daniele Passoni, within his PhD thesis work, in MATLAB® environment, in order to plan the photogrammetric survey considering the expected precision

and, then, to check the quality of the achieved result. It has been integrated in order to manage generic oriented scenarios further than planimetric ones.

Although the implemented tools concern slightly different processing phases, referring to the whole photogrammetric workflow, common threads link all the procedures. Similar strategies have been applied transversely in order to achieve the expected results.

The implementation of MAGO has been a reference in order to approach methods and techniques that have been further applicable even to the other procedures. In particular, the reconstruction of an adaptive mesh in support of the orthophoto generation, as well as the orthogonal projection of the corresponding pixels onto a chosen plane, are being further developed as basic strategies for the prediction of real-time positioning and for the instantaneous 3D scene reconstruction, respectively.

Furthermore, the rotations principles and the point cloud management represent a common ground with respect to U.Ph.O., providing useful prompts to enrich the tool by generalizing the input scenario.

The innovation and the optimization introduced by these instruments are especially useful in such critical cases in which the standard procedure might fail. These scenarios, including, e.g., complex environments or natural disasters sites, are usually characterized by rough operative conditions and/or emergency situations; they have been taken into account even among the analysed case studies.

The implementation of the listed tools, briefly introduced above, is motivated by the novelty of their approaches, compared to the existing ones, in the accomplishment of the same operative steps. Concerning U.Ph.O., there is a wide variety of tools for the mission planning, but it is uncommon that they are able to provide the computation of the expected accuracy, which, however, is a fundamental parameter in determining the quality of the survey.

About the navigation tool, the innovation consists in the addition of the photogrammetric contribution to an integrated real-time navigation and positioning procedure, suitable for indoor environments. An image-based approach lets to overcome the necessity of installing expensive infrastructures, e.g., in critical scenarios. The module for the instantaneous 3D modelling is suitable to obtain a real-time reconstruction, adapting to the changes detected.

Finally, as already stated, MAGO is an innovative tool for the automatic reconstruction of “unrolled” orthophotos of adjacent walls. Its approach is based on the use of the point cloud, instead of the mesh, as input source for the orthophoto reconstruction. This method allows considering each time a specific triangular plane area, on which the image pixel is projected at its original resolution, not predetermined a priori but resulting from the three best-fitting points of the input point cloud. The aim is to avoid the approximations, due to the use of a non-triangular mesh typically obtained by multiple vertices interpolation, as in the best-known software packages.

The dissertation is organized as follows: the first chapter concerns the photogrammetric theoretical principles, together with the state-of-the-art in the photogrammetric software

packages. An overview about the Photogrammetry and CV parallelism is contained in the second chapter, while the third is dedicated to the presentation of the tools. In the fourth chapter, the training test and the validation are shown. Finally, conclusions and future perspectives conclude the thesis.

PHOTOGRAMMETRY: PRINCIPLES AND APPLICATIONS

Photogrammetry is the discipline intended to determine the position and the shape of objects from frames. It represents an image-based 3D survey technique, which starts from at least two shots of the same object, taken from different points of view, in order to reconstruct it in a spatial Reference System (RS). Thus, this technique is based on stereoscopy principles.

Photogrammetry requires the integration by another survey technique, able to acquire the coordinates of the so-called Ground Control Points (GCPs), in order to properly scale and roto-translate the generated model in the chosen RS.

In the past, Photogrammetry was distinguished in terrestrial and aerial; the former was mainly employed in the survey of objects, e.g., buildings, industrial plants, road constructions etc., in order to instantaneously evaluate their state of conservation, while the latter was mainly used in cartography production, together with applications related to the photo-interpretation and classification of land. The UAV Photogrammetry is mostly suitable to cover both the field of application of the other types giving relevant improvements, thanks to the variety of different features achievable by means of the available instruments.

1.1. PHOTOGRAMMETRY ANALYTIC MODELS

The working principles are shown in the following. First of all, a frame is collected; it consists in a central projection of the object points on the image plane (Figure 1-1). When the same point is visible in a second frame, its two projections A' and A" on the different images, called homologous points, form a stereo pair connected to the object point by the homologous rays (Figure 1-2). Once the point is recognizable in at least two frames, the shooting geometry could be restored, making possible the restitution: the 3D positioning of the point is identified by the intersection of the homologous rays.

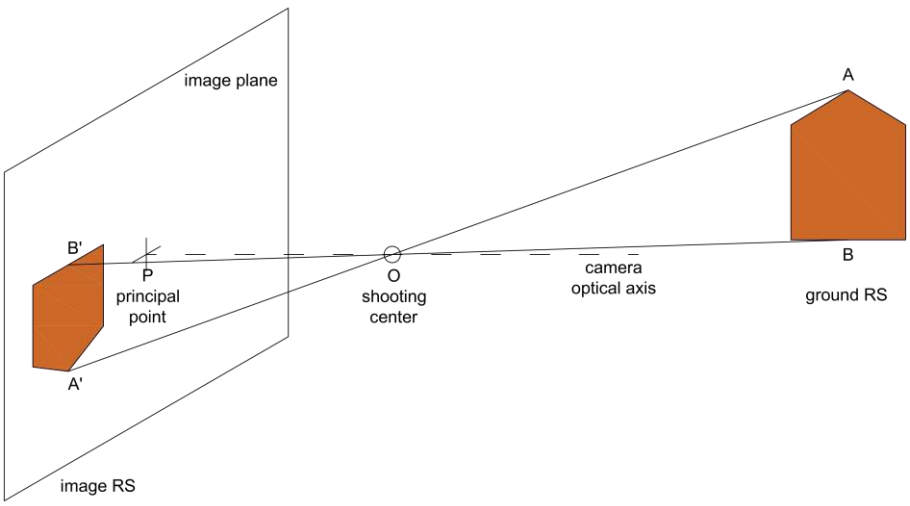


Figure 1-1 Shooting geometry outline.

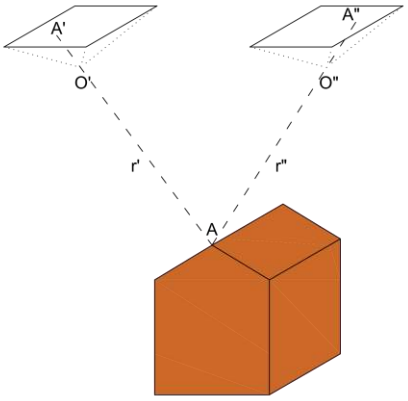


Figure 1-2 Representation of homologous rays, connecting homologous points to the ground point.

The possibility to match homologous points in multiple frames depends on the reciprocal overlapping among the images content. Shooting images along a straight path at a proper distance, called baseline and calculated according to the desired forward overlapping (LO in the following figure) rate, it is possible to obtain a strip. Furthermore, the group of adjacent strips, partially superimposed according to the side overlapping (TO in the following figure) rate, is defined as photogrammetric block. Figure 1-3 depict a schema of the relationships among the single frame, the strip, and the photogrammetric block. The black dots identify the shooting centre of images, while the rectangles correspond to their footprints.

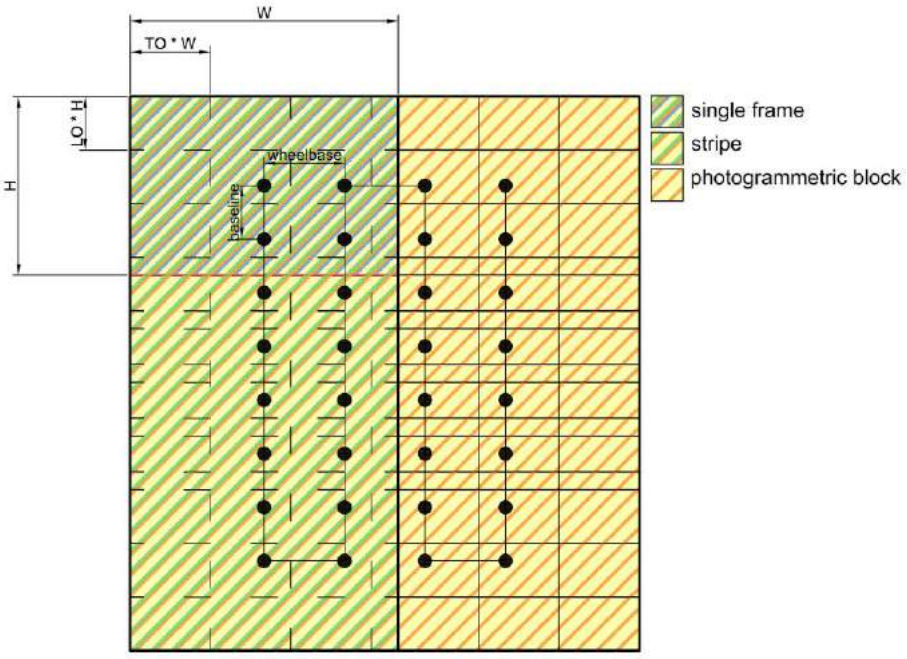


Figure 1-3 Identification of elements within the photogrammetric block.

A photogrammetric campaign is made by the following steps (Cannarozzo et al., 2012):

- *acquisition phase*: the object frames are collected from several positions, making sure that each one partially overlaps with the following and the preceding;
- *orientation phase*: the position and the orientation of images at the time of acquisition are reconstructed, according to mathematical principles;
- *restitution phase*: a 3D model or a point cloud is generated in the three-dimensional space starting from the two-dimensional images.

The geometric relationship between the object point and its projection on the image plane is encoded by the collinearity equations.

$$\begin{aligned} x &= x_0 - c\frac{r_{11}(X_P - X_0) + r_{12}(Y_P - Y_0) + r_{13}(Z_P - Z_0)}{r_{31}(X_P - X_0) + r_{32}(Y_P - Y_0) + r_{33}(Z_P - Z_0)} \\ y &= y_0 - c\frac{r_{21}(X_P - X_0) + r_{22}(Y_P - Y_0) + r_{23}(Z_P - Z_0)}{r_{31}(X_P - X_0) + r_{32}(Y_P - Y_0) + r_{33}(Z_P - Z_0)} \end{aligned} \quad (1.1)$$

In equation (1.1), the involved parameters are represented by: the point coordinates $\underline{X}_P = (X_P, Y_P, Z_P)$, the image coordinates (x, y), the Internal Orientation (IO) parameters (focal length c, and principal point coordinates, x_0, y_0), the External Orientation (EO) parameters (camera positions $\underline{X}_0 = (X_0, Y_0, Z_0)$ and r_{ij} components of the Cardano rotation matrix $\boldsymbol{R}$). In Table 1-1, the list of the basic operations associated with the searched unknown is reported.

Table 1-1 Photogrammetric parameters and their applications.

Known	**Obtained**	**Operation**
IO, EO, $\underline{X}_P$	(x, y)	shot simulation
IO, $\underline{X}_P$, (x, y)	EO	external orientation
[IO, EO, (x, y)] × N > 1	$\underline{X}_P$	restitution
IO, EO, $\underline{X}_{PG}$, (x, y)	$\underline{X}_{PC}$	orthophoto

where N represents the number of images framing the same object point, while $\underline{X}_{PG}$ and $\underline{X}_{PC}$ are the object point geometry and colour respectively.

It is worth noting that, in the real practice, more than one of these operations is required at the same time. For this reason, a least squares approach, called Bundle Block Adjustment (BBA), is used to properly estimate the unknown parameters. Bundle Block Adjustment takes

into account the direct relationship between image and ground, consisting in collinearity equations.

The workflow starts matching tie points, corresponding to the notion of homologous points previously mentioned, within the frames of the photogrammetric block. This operation takes into account their common features, hence these points need to be univocally discerned. Thus, their image coordinates are measured in the images of the block. Using the collinearity equations, the image coordinates, the corresponding ground coordinates, the IO and EO parameters are related. Thus, by means of a simultaneous least squares adjustment, the ground coordinates of the tie points, the EO and IO parameters could be estimated.

The image coordinate measurements and the IO parameters define a bundle of light rays, while the EO parameters define the position and the attitude of the bundles in space. During the adjustment the bundles are rotated (ω, ϕ, κ) and shifted (X_0, Y_0, Z_0) until: the conjugate light rays intersect as well as possible at the locations of object space tie points; light rays corresponding to GCPs pass through the object points as close as possible.

Unknown parameters might include:

- ground coordinates of tie points;
- EO parameters of the involved frames;
- IO parameters of the involved cameras.

Observable quantities might include:

- ground coordinates of control points;
- image coordinates of tie/control points;
- IO parameters of the involved cameras, if they have been calibrated in advance;
- EO parameters of the involved frames, e.g., from a GPS/INS unit on-board.

Since it is a non-linear problem, a preliminary step for linearization is required.

The described approach, at the basis of the best-known photogrammetric software packages, is based on the simultaneous bundle block adjustment. It is a rigorous method that, however, does not take advantage of stereovision.

A similar approach consists in the *projective stars* model, which aims solving simultaneously all the collinearity equations of the photogrammetric block, considering as observables the image coordinates of the tie/control points and assuming as known the internal orientation parameters and the object coordinates of the control points. The unknowns are represented by the external orientation parameters of the shooting centres, together with the ground coordinates of the tie points.

The alternative to these methods is represented by the *independent models* approach, which follows a step-by-step procedure. The first phase is the estimation of the Relative Orientation (RO) of an image with respect to another, using a stereoscopic model (u, v, w); in this way, the shooting geometry is recovered and the object is put in a model RS. Then, the Absolute Orientation (AO) is computed thanks to the GCPs, since both their image and ground

coordinates are known; by means of a roto-translation and scale variation, it is possible to pass from the model RS to the external RS. At least five pairs of homologous points are needed to recover the RO of frames; in facts, each camera has six degrees of freedom in space, corresponding to a total of 12 for a couple of images. Nevertheless, the conditions are invariant for three translations, three rotations and a scale factor, leading to five degrees of freedom depending on the co-planarity equations (due to epipolar geometry), which relates the homologous rays connecting the homologous points. The seven remaining degrees of freedom, which determine the external RS, are locked through the GCPs coordinates, at least two with both planimetric and altimetric information and another with just the altimetric one. The recovered stereovision is exploited in the collimation of the wished object point, and the final AO is computed according to the following relation:

$$\begin{bmatrix} X_P \\ Y_P \\ Z_P \end{bmatrix} = \lambda R \begin{bmatrix} u \\ v \\ w \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \tag{1.2}$$

This technique was commonly used in restitution by means of analogic instrumentation.

Comparing the two approaches (Ackermann, 1973; Gosh, 2005; Gomarasca, 2009), on the one hand the main pro of Bundle Block Adjustment is due to the possibility of fixing errors without repeating all the procedure, while in independent models the workflow need to be restored starting from the step affected by the error. On the other hand, the former method requires a long time to let the operator collimate all the recurrences of the points in the images, while in the latter a single collimation allows identifying homologous points onto two frames at time, exploiting stereovision.

Bundle Block Adjustment method is at the basis of Structure From Motion (SFM; Ullman, 1979) technique, implemented in the main photogrammetric software packages. The cons of this model, due to the burden of collimating each single recurrence of homologous points in the frames, is solved thanks to the high level of automation achieved for the contribution of Computer Vision (CV).

1.1.1. Image and ground Reference Systems

In the previous equation (1.1) the expression of the collinearity ray is reported; it links the ground point with its projection on the image. The relationship between the internal RS of the image x^* and the external one of the object X is contained in the following expression.

$$\begin{bmatrix} x^* \\ y^* \\ z^* \end{bmatrix} = \lambda R_C \begin{bmatrix} X_P - X_0 \\ Y_P - Y_0 \\ Z_P - Z_0 \end{bmatrix} \tag{1.3}$$

The rotation matrix R_C used in the equation (1.3) is built according to the Cardano convention, typically employed in Photogrammetry; thus, it is computed considering the rotation angles with respect to the axes progressively rotated.

$$R_C = R_\kappa \cdot R_\varphi \cdot R_\omega \tag{1.4}$$

The elements inside the matrices are:

$$R_\kappa = \begin{bmatrix} \cos\kappa & \sin\kappa & 0 \\ -\sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.5}$$

$$R_\varphi = \begin{bmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{bmatrix} \tag{1.6}$$

$$R_\omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \sin\omega \\ 0 & -\sin\omega & \cos\omega \end{bmatrix} \tag{1.7}$$

Thus, the single entries of the Cardano rotation matrix are:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

$$\begin{aligned}
r_{11} &= \cos\varphi \cos\kappa \\
r_{12} &= \cos\omega \sin\kappa + \sin\omega \sin\varphi \cos\kappa \\
r_{13} &= \sin\omega \sin\kappa - \cos\omega \sin\varphi \cos\kappa \\
r_{21} &= -\cos\varphi \sin\kappa \\
r_{22} &= \cos\omega \cos\kappa - \sin\omega \sin\varphi \sin\kappa \\
r_{23} &= \sin\omega \cos\kappa + \cos\omega \sin\varphi \sin\kappa \\
r_{31} &= \sin\varphi \\
r_{32} &= -\sin\omega \cos\varphi \\
r_{33} &= \cos\omega \cos\varphi
\end{aligned} \tag{1.8}$$

On the contrary, using the function $atan2(y, x)$ on the elements of R_C matrix, it is possible to obtain ω, φ and κ angles as:

$$\begin{aligned}\omega &= \text{atan2}(-r_{32}; r_{33}) \\ \varphi &= \text{atan2}(r_{31}; \sqrt{r_{32}^2+r_{33}^2}) \\ \kappa &= \text{atan2}(-r_{21}; r_{11})\end{aligned} \tag{1.9}$$

Another admitted solution is represented by:

$$R(\kappa)R(\varphi)R(\omega)=R(\kappa+\pi)R(\pi-\varphi)R(\omega+\pi) \tag{1.10}$$

The so-obtained rotation matrix R_C allows to relate the auxiliary RSs, both centred in the same origin O and parallel to the image and ground RS respectively. The point O represent the position of the shooting centre of the camera, called principal point.

1.1.2. Resume about the main rotation conventions

Rotation matrices are characterized by two main properties. They are orthogonal, so that:

$$RR^T=R^TR=I \tag{1.11}$$

and their determinant is equal to 1; otherwise, if their determinant is equal to -1, they are not proper rotation matrices but roto-reflection matrices.

A rotation matrix has nine elements, which represent the constraints of the spatial rotations (three degrees of freedom) and the ones related to the right-handed coordinate system (six degrees of freedom).

There are several conventions to represent rotations; in photogrammetry, the Cardano angles presented in paragraph 1.1.1 are mainly used.

Other possible representation among the best-known are: Rodrigues rotation matrix, Euler angles, axis-angle rotation, quaternions (Förstner and Wrobel, 2016).

The Rodrigues rotation matrix (Rodrigues, 1840) represent a formulation by analytical description, as reported in the following:

$$R(\theta)=e^{S(\theta)}=I_3+\frac{\sin\theta}{\theta}S_\theta+\frac{1-\cos\theta}{\theta^2}S_\theta^2 \tag{1.12}$$

with $\theta = |\theta|$. The vector θ parametrize the rotation, while the $S_\theta = lnR(\theta)$ represents the so-called skew matrix.

Concerning the Euler angles (Euler, 1770; Euler, 1775; Kanatani, 1990), called α, β and γ, they represent the single rotations of the new triplet of Cartesian axes with respect to the

starting axes XYZ. Different concatenation rules could be applied to obtain general rotations from the elementary rotations $R(\alpha)$, $R(\beta)$ and $R(\gamma)$.

$$R(\alpha)=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\alpha & -\sin\alpha\\ 0 & \sin\alpha & \cos\alpha\end{bmatrix} \tag{1.13}$$

$$R(\beta)=\begin{bmatrix}\cos\beta & 0 & \sin\beta\\ 0 & 1 & 0\\ -\sin\beta & 0 & \cos\beta\end{bmatrix} \tag{1.14}$$

$$R(\gamma)=\begin{bmatrix}\cos\gamma & -\sin\gamma & 0\\ \sin\gamma & \cos\gamma & 0\\ 0 & 0 & 1\end{bmatrix} \tag{1.15}$$

The so-called method axis-angle is gathered from the Rodrigues rotation.

$$R_{r,\theta}=I_3+\sin\theta\, S_r+(1-\cos\theta)S_r^2 \tag{1.16}$$

where r is the rotation axis and represent a normalized vector with module equal to 1.

Concerning the quaternions (Hamilton, 1844), they are homogeneous 4-vectors represented as:

$$q=\begin{bmatrix}q_0\\ q_1\\ q_2\\ q_3\end{bmatrix}=\begin{bmatrix}q\\ \mathbf{q}\end{bmatrix} \tag{1.17}$$

They consist in a scalar part $q = q_0$ and an imaginary part behaving like a vector $\boldsymbol{q} = [q_1, q_2, q_3]^T$. If a quaternion is treated as a vector, the first element, the scalar part, is denoted by q_0; otherwise, if it is treated as an algebraic entity, the scalar part is denoted by q.

Considering a unit vector $\bar{r}$ as rotation axis and the quaternion $q = \cos\frac{\alpha}{2} + \bar{r}\, sin\frac{\alpha}{2}$, the operation:

$$\bar{p}'=q\bar{p}q^{-1} \tag{1.18}$$

yields the vector $\bar{p}$ rotated by an angle α around the axis $\bar{r}$.

The equivalent representation in terms of axis-angle rotation matrix is obtainable from (1.16) as:

$$R_Q = \frac{1}{|\mathbf{q}|^2}((q^2 - q^T q) I_3 + 2D_q + 2qS_q \quad (1.19)$$

1.1.3. Photogrammetric block geometries and survey planning

The typical shooting geometry of aerial Photogrammetry consists of a sequence of forward overlapping frames, called strip, acquired with a nadir attitude. The set of strips, overlapping even in the transversal direction, is called photogrammetric block, as previously presented in Figure 1-3.

UAV Photogrammetry inherited this shooting geometry, together with the one mostly typical of terrestrial Photogrammetry, i.e., convergent on the surveyed object. In this case, the strip is taken surrounding the object, without neglecting the adequate overlapping among images.

The choice of the shooting geometry depends on the aim, the context in which the survey is performed, and the shape of the object, as well as the available instrumentation. The inputs involved in the campaign planning are the camera parameters (focal distance *c* and the size of image, frame and pixel), the site dimensions, the overlapping percentages among frames, the flight height, and the shooting time interval (in case of automatic snapshots). Once these options are set, it is possible to infer the survey resolution and precision, together with the number of collected images, the distances among them (baseline and wheelbase) and the time needed for the campaign, checking if they are satisfactory.

In the digital Photogrammetry, the Ground Sample Distance (GSD), i.e., the size of the projection of the single pixel on the ground, is the maximum resolution, which is typically assimilated to the graphic error in the graphical representations. In both cases, GSD and graphic error are the highest level of detail and reliability reachable.

In Table 1-2, the equations leading to the relevant parameters inferable during the planning phase, i.e., scale factor *n*, GSD, FOV (Field Of View), longitudinal baseline *B*, transversal wheelbase *i*, altimetric precision σ_Z (depending on the accuracy of the collimations on images $\sigma_{p\xi}$), and number of frames, are reported, together with the input parameters.

Table 1-2 Planning parameters, divided in inputs and resulting checking parameters.

	Inputs			Checking parameters		
CAMERA PARAMETERS	focal length	c	[mm]	GSD	$n \cdot d$	[m]
	width (frame size)	W_f	[mm]	long. baseline	$B = (1\text{-}LO) \cdot H_f \cdot n$	[m]
	height (frame size)	H_f	[mm]	trans. wheelbase	$i = (1\text{-}TO) \cdot W_f \cdot n$	[m]
	width (image size)	W_i	[pixel]	σ_Z	$Z^2/cB \cdot \sigma_{p\xi}$	[m]
	height (image size)	H_i	[pixel]	number of frames	$n_f = (1 + A_l/B) \cdot (1 + A_t/i)$	[-]
	pixel size	d	[μm]	FOV	$(W_f \cdot n) \times (H_f \cdot n)$	[m^2]
GEOMETRY	area longitudinal side	A_l	[m]	time needed	$t \cdot n_f$	[s]
	area transversal side	A_t	[m]			
	relative flight height	Z	[m]			
	long. overlapping	LO	[%]			
	trans. overlapping	TO	[%]			
	resulting scale factor	n = Z/c	[-]			
	shooting time interval	t	[s]			

Considering a regular sequence of frames, the forward and side overlapping are chosen on the basis of the type of lens and the presence of obstructions. Shorter focal distances allow a larger FOV, but they are more sensitive to obstructions.

Another criterion in the choice of the overlapping percentages is the method employed for the restitution phase. In facts, using the model commonly implemented in the photogrammetric software packages by means of the BBA algorithm, it is convenient if a single point appears in more than two images, in order to increase the number of observations and, hence, the constraints of the whole block, which is examined in its entirety. For this reason, both the forward and side overlapping are usually set to values around the 70-80%.

Considering the stereo pairs used in the *independent model* approach, a lower percentage of overlapping is required, typically the 60% in longitudinal direction and the 30% in the transversal one, since just the couple of adjacent frames is taken into account in the step-by-step procedure. Moreover, the short length of the baseline adversely conditions the expected precision according to the classical Photogrammetry approach (Krauss, 1993), as inferable from equation (1.20).

$$\sigma_Z = \frac{Z^2}{cB}\sigma_{p\xi} \tag{1.20}$$

1.1.4. Comparison with Laser Scanner pros and cons

The photogrammetric technique, described in the previous paragraphs, is suitable for a wide range of applications. In the majority of the case studies exploited in this thesis work, Photogrammetry has been used combined with UAV. In facts, the flexibility resulting from the use of cameras mounted on drones enlarges the range of possible scenarios, including complex environments and conditions. Furthermore, UAV Photogrammetry is cheap, fast and reliable, allowing the accessibility even in rough areas without compromising the operators' safety.

Compared to the main competitor, i.e., Laser Scanner (LS), some pros and cons are noticeable. The necessity of another supporting technique to perform the survey of the GCPs coordinates, in order to scale and roto-translate the model in the proper RS, as well as the long time needed for the post-processing are disadvantageous aspects for Photogrammetry; conversely, LS acquisitions are already correctly scaled and require a shorter time to be reciprocally registered. However, Photogrammetry accounts for a cheaper instrumentation and allows the collection of the colour information.

In recent times, the LS technique has been significantly improved, taking inspiration from the peculiarities and the goals of Photogrammetry. Two main contributions could be highlighted: on the one hand, some instruments have been integrated with cameras to associate the colour information to the LS point cloud; on the other hand, the LS size and weight have been substantially reduced to allow to be mounted on drones, as well as cameras.

These circumstances underline the importance of the integration of techniques, to gain benefits from their different main strengths.

1.2. DEVICES

There is a wide spectrum of instruments available for Photogrammetry; potentially, just a camera is needed and many devices have it embedded. From the professional point of view, the sensor type and the readout method are the main aspects that influence the quality of the instrumentation.

The sensor type could be distinguished in CMOS (Complementary Metal Oxide Semiconductor) and CCD (Charge-Coupled Device) (Table 1-3). In a CCD sensor, every pixel's charge is transferred through a very limited number of output nodes, often just one, to be converted to voltage, buffered, and sent off-chip as an analogue signal. All the pixels can be devoted to light capture, and the output's uniformity is high. In a CMOS sensor, each pixel has its own charge-to-voltage conversion, and the sensor often includes amplifiers, noise-correction, and digitization circuits, so that the chip outputs digital bits.

CMOS technology, which was born a few years later and with lower performances with respect to the CCD, has been significantly enhanced in the last years, making the two alternatives substantially comparable. CMOS is nowadays the most spread, thanks to its features of high speed (in terms of frame rates), high resolution (number of pixels), strong dynamic performance, low power consumption, improved noise performance, improved quantum efficiency, improved colour concepts, good price/performance ratio. Nevertheless, CCD is still more high-performance in terms of noise affliction because of the different constructive principles.

Table 1-3 CMOS vs CCD sensors: basic features.

Features	**CCD**	**CMOS**
Photodiode output	electric charge	voltage
Chip output	voltage (analogic)	bit (digital)
Camera output	bit (digital)	bit (digital)
Noise	low	moderate
Sensor complexity	low	high
Dynamic range	wide	moderate
Uniformity	high	low to moderate
Burst speed	moderate to high	high
Chromatic precision	high	medium

Concerning the readout method, in the so-called Global Shutter approach all the pixels of the array are exposed simultaneously, thus, enabling freeze frame capture of fast moving or fast changing events. Global Shutter mode is very simple to synchronize and often yields faster frame rates. A reference readout is performed in addition to the actual readout of charge from each pixel, in order to eliminate reset noise from the image.

Global Shutter is counterpoised to Rolling Shutter method, in which the acquisition of the light impulse is not simultaneous in all the pixel of the sensor. This property makes the Rolling Shutter method more sensitive to the dragging effects.

In the present work, the performed tests have been mainly achieved by using UAV Photogrammetry. Both multi-rotor and fixed wing drones have been employed. The former are more suitable in case of heterogeneous scenarios, e.g., with height surges, and small areas,

thanks to their higher freedom of movement; the latter glide, thus they are ideal for huge and flat areas.

From an operational point of view, the ready-made UAVs have been equipped with their own embedded cameras; conversely, self-built drones have been equipped with Reflex cameras.

Besides the cited instrumentation, even high-performance cameras, certified also for military applications, have been introduced to be employed in two specular scenarios requiring a high Frame (Rate) Per Second (FPS). In the former situation a couple of cameras, mounted on a car rack, move with the vehicle in order to survey the surrounding environment, while in the latter the cameras are in a fixed position allowing the 3D reconstruction of the changing scene.

A mention is owed to fisheye cameras, although they have not been used in anyone of the performed experimentations. In facts, the interest of the scientific community about their employment in the photogrammetric field is increasing, e.g., in solving indoor architectural and positioning issues (Perfetti et al., 2018; Tomaselli et al., 2019), thanks to their low cost and large field of view.

1.2.1. Camera calibration

The cameras employed in the surveys are not geometrically perfect. Thus, the measurements taken from the images are affected by the contribution of distortions.

In order to estimate and remove them, the camera calibration procedure allows to recover the IO parameters, including the aforementioned principal distance c and principal point coordinates (ξ_0, η_0), together with the radial (K_1, K_2, K_3), tangential (P_1, P_2) and skew (b_1, b_2) distortions characterizing the lens.

$$r=\sqrt{(\xi-\xi_0)^2+(\eta-\eta_0)^2} \tag{1.21}$$

$$d\xi=(\xi-\xi_0)[K_1\cdot r^2+K_2\cdot r^4+K_3\cdot r^6]+P_1[r^2+2(\xi-\xi_0)^2]+2P_2(\xi-\xi_0)(\eta-\eta_0)+ +b_1(\eta-\eta_0)(\xi-\xi_0)+b_2(\eta-\eta_0) \tag{1.22}$$

$$d\eta=-(\eta_0)[K_1\cdot r^2+K_2\cdot r^4+K_3\cdot r^6]+P_2[r^2+2(\eta-\eta_0)^2]+2P_1(\xi-\xi_0)(\eta-\eta_0) \tag{1.23}$$

$d\xi$ and $d\eta$ represent the correction of the image coordinates, in x and y axes of the image plane respectively, due to lens distortions, while r is the radial distance from the generic point on the image plane (ξ,η) to the principal point.

Camera calibration could be achieved according to different techniques:

- *laboratory calibration*: it is extremely precise, but very expensive, since it requires the use of instruments such as collimators or goniometers;

- *calibration with a 3D test field*: it is achievable by acquiring a sequence of images of a 3D object characterized by a known geometry;
- *calibration with a 2D test field*: it is achievable by acquiring a sequence of images of a flat panel characterized by a known geometry;
- *simultaneous calibration*: by means of BBA, a self-calibration is performed by simultaneously estimating the 3D coordinates of the object and the IO parameters;
- *vanishing point method*: only the coordinates of the principal point are estimated.

1.3. POST-PROCESSING APPLICATIONS

For sake of simplicity, the present paragraph refers to a specific case study, in order to illustrate the features of a selection of photogrammetric software packages.

The photogrammetric post-processing consists in several steps allowing to obtain the final outputs. The achievable products are dense cloud, mesh, Digital Surface Model (DSM), and orthophoto. Dense cloud and mesh (Figure 1-4) are 3D entities, made by points and polygons respectively, while the DSM is a raster image constituted by cells, containing information on the elevation concerning both the natural and artificial features of the environment, and the orthophoto consists in an orthogonal projection of the object or scene on a defined plane (Figure 1-5).

Figure 1-4 An example of photogrammetric dense point cloud (left) and the correspondent textured meshed model (right).

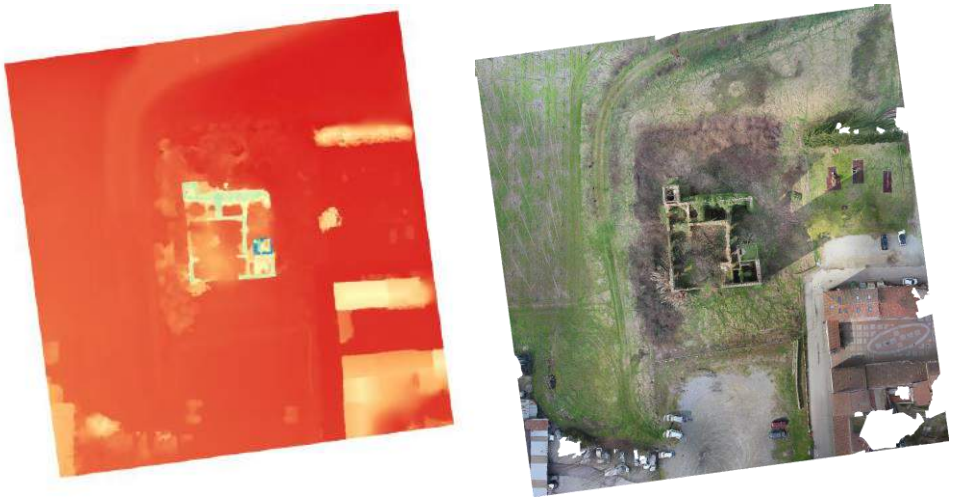


Figure 1-5 An example of DSM (left) and correspondent orthophoto (right).

A multiplicity of free and commercial software packages are available to efficiently perform the photogrammetric workflow. In the following paragraph, a selection of the best-known and wide-spread ones is presented.

1.3.1. Software comparison

The innovations in the field of Photogrammetry, concerning matching algorithms and quality of digital airborne cameras, have led to a substantial increase of the quality of 3D models obtained by images. Hence, the theme of photogrammetric software comparison is a topic of interest in the scientific literature. Different aspects related to a wide selection of case studies have been treated, including whether urban (Haala and Cavegn, 2016), historical (Alidoost and Arefi, 2017; Pavelka et al., 2013; Wojtas, 2010) and environmental scenarios (Burns and Delparte, 2017; Svensk, 2017; Niederheiser et al., 2016; Gini et al., 2013) or focusing the attention on specific outputs and parameters (Gabara and Sawicki, 2017; Schöning and Heidemann, 2015; Remondino et al., 2014; Haala, 2013).

A group of programs has been compared basing on the performances and the output quality in Gagliolo et al. (2018b), taking into account as case study the Castle of Casalbagliano in Alessandria (Italy). A particular focus has been addressed to the Agisoft Photoscan© (now called Metashape©; Agisoft© LLC, 2019) workflow even in Gagliolo et al. (2017 and 2018a), since it has been mostly employed during the experimental activities thanks to its user-friendliness. The list of tested software includes Agisoft PhotoScan© v. 1.2.6, Pix4Dmapper© v. 4.1.25 (Pix4D SA, 2019), ContextCapture™ v. 4.4.7.68 (Bentley Systems Incorporated, 2019), VisualSFM v. 0.5.25 (Wu, 2013; Wu et al., 2011), insight3d 0.3.2 (Mach, 2009) and MicMac v. 1.0.beta11 (IGN and ENSG, 2019).

These photogrammetric software have been analysed and evaluated, paying particular attention to the following features: (1) the possibility to integrate several datasets of images taken by different points of view and eventually with different focal distances, (2) the behaviour in presence of GCPs, (3) the speed of calculation, (4) the user-friendliness and

completeness of the workflow, (5) the interoperability with external software and platforms, (6) the entity of user communities, (7) the availability of documentation, tutorials and direct support, and (8) the cost.

The dataset employed for the comparison includes three different shooting geometries, nadir (N, 83 frames), oblique (O, 61 frames) and terrestrial (T, 28 frames), acquired using the Canon EOS-M camera, with a fixed focal distance of 22 mm. The nadir and oblique shooting geometries have been obtained thanks to the use of the self-built UAV Microcopter hexacopter; the relative camera locations are shown in Figure 1-6.

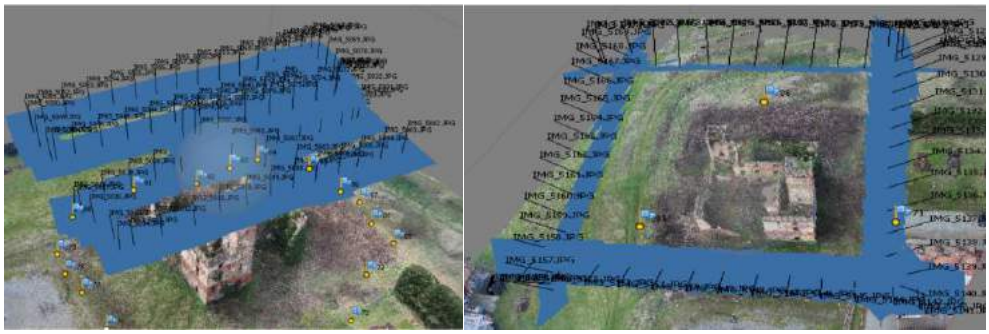


Figure 1-6. Camera locations for the nadir (left) and oblique (right) photogrammetric blocks from Agisoft Photoscan©

The surveyed area is a square with a side of about 60 m, with the Castle located in the centre; the resulting Ground Sample Distance (GSD) is about 0.01 m. The images have been georeferenced by means of 19 square targets, whose coordinates have been surveyed with the Total Station (TS) in a local Reference Frame (RF). The used TS, i.e., Leica Nova Multistation MS60, is characterized by an angular accuracy of 3^{cc} and by a distance accuracy of 1 mm + 1.5 ppm using reflective prism (Fagandini et al., 2017).

The ground truth considered to make the output comparison and validation homogeneous is a Terrestrial Laser Scanner (TLS) point cloud of the same area, acquired by means of a Z+F Imager® 5006h executing and registering eight scans. The scans positions have been located around the whole structure at a distance of about 30-40 m from it. A resolution setting of 20,000 points/360° has been used to guarantee a spacing of about 0.01 m in the acquired point clouds. Additional 32 TLS targets have been surveyed to allow the georeferencing in the same local RF of the photogrammetric point clouds.

Although it is possible to fully reconstruct the whole structure only using the O dataset, the ensemble of images (T + N + O) has been processed by the tested software in order to analyse their behaviour dealing with different attitudes. Instead, only the O dataset has been used for the complete workflow, i.e., aerotriangulation (obtaining IO and EO from BBA), dense cloud generation, mesh reconstruction, orthophoto production. The N and T images have not been included in the complete processing to speed up the elaboration times; moreover, they do not

give individually a complete view of the object, because they focus on a preferential portion consisting in roofs and façades, respectively.

All the phases of the workflow are performed using a *medium quality*, corresponding to a downsampling factor of 4 with respect to the original images. The analysis discussed have been performed using a desktop Personal Computer (64 bits Windows 7 operating system, Intel® core™ i7-4770 CPU @340 GHz processor and 32 GB of RAM, Intel® HD Graphics 4600, 20 Cores @400 MHz, 1297 MB).

The main features of the analysed photogrammetric software concerning the post-processing are summarized and compared in Table 1-4.

Table 1-4 Summary report about software features.

	Agisoft PhotoScan©	**Pix4D©**	**ContextCapture™**	**VisualSFM**	**MicMac**	**insight3d**
License	P	P	P	F	F - OS	F - OS
Time	○	>	<<<	<<	<	✗
Integration of datasets	✓	✗	✓	✓	✓	✗
GCPs input	✓	✓	✓	✓	✓	✗
User-friendliness	○	=	=	<	<<	<
Workflow	○	=	=	<	=	<<<
Interoperability	○	<	<	<	<	<<<
User communities	○	=	<	<	=	<<<
Documentation, tutorials and direct support	○	=	=	<<	=	<<<
Cost (compared with Agisoft Photoscan©)	n	2n	>2n	0	0	0

Legend
P: Proprietary; F: Free; OS: Open Source;
○ : used as reference; ✓: present; ✗: absent;
<, <<, <<<: minor (increasing levels);
>, >>, >>>: major (increasing levels);
=: equal.

A TLS point cloud is used as reference in the comparison with the photogrammetric ones, thanks to the favourable conditions of the TLS survey campaign that allowed producing a

reliable result. The use of TLS as an independent reference makes the comparisons homogeneous and coherent.

The East side of the tower (in particular the yellow rectangle in Figure 1-7) has been chosen as representative portion for the comparison, because it clearly appears in the majority of the images. The TLS noise measurement on this portion, with a low reflectivity and a distance of 30 m, amounts to about 2 mm RMS.

A first comparison is performed evaluating the point clouds density on a sample of one square meter, located inside that portion of the tower (in red in Figure 1-7).

Then, the signed distances between two point clouds is computed by M3C2 (Multiscale Model to Model Cloud Comparison) method implemented in CloudCompare (CloudCompare Development Team, 2021; Lague et al., 2013).

Each point cloud density has been evaluated in terms of points per m^2, considering the red sample shown in Figure 1-7. The results are reported in Table 1-5.

Table 1-5 Points per square meter in the evaluated sample.

Source	**Points per m^2 [-]**	**Spacing b/w points [mm]**
TLS	12202	9
Agisoft PhotoScan©	552	43
Pix4D©	1556	25
ContextCapture™	1976	22
VisualSFM	1097	30
MicMac	1446	26

Figure 1-7. In yellow, the sample used for M3C2 comparison; in red the sample of a square meter, used to compute the density.

It is evident from Table 1-5 that the density of the cloud resulting from TLS survey is much higher than the one of the photogrammetric products. Although the processing parameters have been chosen in order to achieve homogeneous conditions, the Agisoft Photoscan© point cloud is much sparser than other software packages, with a resulting number of points per m^2 from two to four times less than the others and, consequently, a nearly doubled spacing between points.

Subsequently, M3C2 plugin computes signed distances between two point clouds. It has been employed setting the whole TLS point cloud as reference cloud (Cloud#1), without subsampling; thus, the *core points*, on which the computation is performed, include all the points of the cloud. The preferred orientation of point cloud perpendicular lines has been set on "+X", considering the East side of the tower almost parallel to Y axis. In the *output* section, the projection of the *core points* is done keeping original positions, i.e., a new cloud containing the differences between the tested point clouds is generated.

Once the M3C2 algorithm has been applied, the resulting cloud has been analysed with the support of the scalar fields (SF) display parameters tool, in order to allow the customization of its visualization according to the most suitable legend. The differences have been represented fitting a Gaussian distribution to them (Figure 1-8), cutting off the tails and setting the evaluation range from -0.1 to +0.1 m.

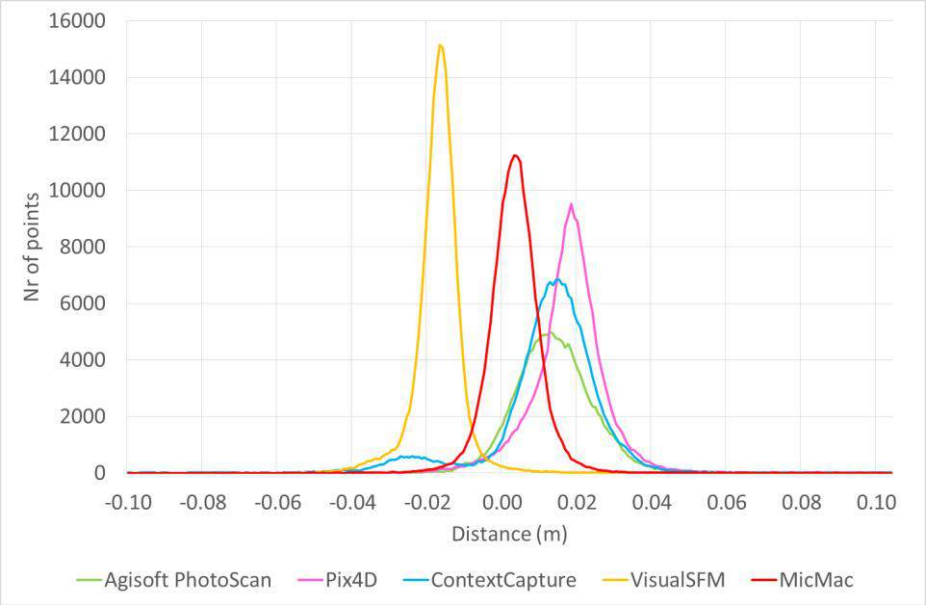


Figure 1-8. Results of M3C2 comparisons.

All the solutions except for MicMac show a remarkable bias with respect to the TLS point cloud, thus a difference with respect to the expected mean value corresponding to 0 in case of well overlapped clouds: about +0.01 m for Agisoft PhotoScan© and ContextCapture™, +0.015 m for Pix4D©, and -0.02m for VisualSFM. This could be connected with the RF. In fact, the maximum observed shift is approximately twice the GSD.

The dense clouds from Agisoft PhotoScan© and MicMac have been georeferenced using the GCPs picking in the software themselves; the quality of this operation is assessable by means of the Root Mean Square Error (RMSE) provided in the software interface itself, consisting in the square root of the quadratic mean of differences between values predicted by a hypothetical model and the observed values. Pix4D© cloud has been referenced in UTM, because the software doesn't work easily in a local RF. A further roto-translation has been applied, by giving the new coordinates of the GCPs. Finally, ContextCapture™ and VisualSFM clouds have been aligned (roto-translation and scale) by means of CloudCompare.

If the objective is just the survey of the structure in a local RS, as in the analysed case, the bias of the mean value is not significant. On the other hand, if the building must be seen in a global RF, this error has to be taken into account and evaluated in the light of the representation scale needed.

Nevertheless, the comparisons highlight a good behaviour for all the examined software, also inferable from the low values of standard deviation, included in the range of 0.007 - 0.015 m (Table 1-6). In facts, this parameter is more representative of the overall quality of the clouds, giving the size of the range of the differences with respect of TLS while disregarding the eventual bias of average value maybe due to the different attribution of RF.

Table 1-6. M3C2 results in terms of Gaussian mean and standard deviation of the differences.

	Mean [m]	**Std [m]**
Agisoft PhotoScan©	0.012	0.010
Pix4D©	0.015	0.010
ContextCapture™	0.010	0.015
VisualSFM	-0.019	0.007
MicMac	0.001	0.007

In Figure 1-9, the maps of the M3C2 distances, obtained comparing TLS with all the single clouds, are shown. In general, the worst portions are located right by the holes of the masonry and the crenulation (located around 3/4 of the tower top), because of the difficulty of their reconstruction by the photogrammetric software. In cases of VisualSFM and MicMac, according to the lower standard deviation, the distribution is more homogeneous than other software. The maps related to Agisoft PhotoScan© and Pix4D© are characterized by a "diffuse dirt", while ContextCapture™ one enhances localized areas with higher standard deviations, mainly around the crenulation, which represents the only dept discontinuity of the chosen sample. Figure 1-9 suggests that the first and the second software packages allow producing an homogeneous but poorer result than the third one, whose lacks in accuracy are focused around the discontinuity portion.

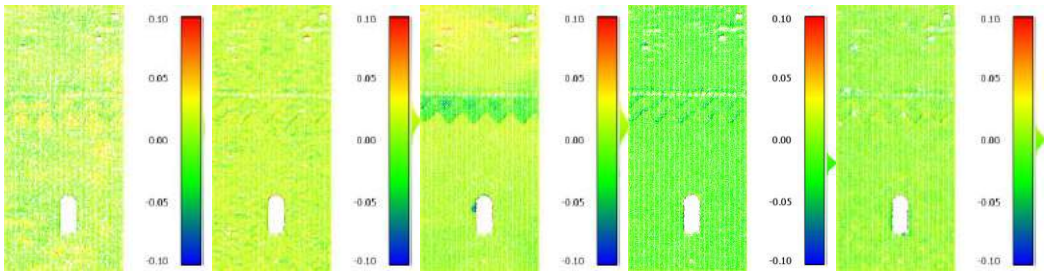


Figure 1-9 Map of the M3C2 distances between TLS and (from left to right): Agisoft PhotoScan©, Pix4D©, ContextCapture™, VisualSFM and MicMac.

In conclusion, the aim of this software comparison is not the establishment of a ranking, but the underlining of the potentialities of the examined packages, in order to be more aware in the choice of the most appropriate tool for the specific scenario of interest.

COMPUTER VISION FOR AUTOMATION IN PHOTOGRAMMETRY

As already stated, the most common software packages, including the aforementioned ones, work on the basis of the BBA algorithm, described in Section 1.1. This approach, as well as the whole photogrammetric workflow, is strongly influenced by the integration of Photogrammetry with Computer Vision (CV), which gives rise to the so-called Structure From Motion (SFM) technique (Forsyth and Ponce, 2012; Förstner and Wrobel, 2016; Triggs et al., 2000). Computer Vision is an innovative scientific field, related to the Artificial Intelligence (AI) and the Machine Learning (ML), which deals with robotics scene understanding, starting from the interaction based on images and/or videos acquisition. In particular, it aims achieving those tasks performed by biological vision systems, e.g., seeing or sensing a visual stimulus, understanding what is being seen, and extracting complex information in order to use it in further processes. The goals of this interdisciplinary field are the simulation and automation of such tasks of human vision systems by using sensors, computers, and machine learning algorithms. In summary, CV is the theory underlying artificial intelligence systems' ability to see and understand their surrounding environment.

Some of the most popular fields of application of this discipline concern: autonomous vehicles and robots, needing to gather information about their surroundings to decide how to behave; facial recognition, in order to check the personal identity and to guarantee the access to confidential data; image search and object recognition, identifying objects within images and extracting information.

Since both Photogrammetry and Computer Vision aim to obtain 3D information from images, the CV principles and techniques are deeply integrated in the photogrammetric software processing, thanks to their significant contribution in the automation of operations. The integration of CV in Photogrammetry (Barazzetti et al., 2009; Qin and Grün, 2020) is noticeable in the whole workflow from data acquisition, geometric data processing, 2D and 3D interpretation, recognition modeling, to data administration and representation (Grün et al., 2009; Lu et al., 2004; Qin, 2015; Qin et al., 2016; Shan et al., 2020). In particular, the development of more automated methods involving the use of AI techniques aim to both automate processes that traditionally require heavy manual operation, such as modeling (Grün and Wang, 1998; Lillesand et al., 2014), and improve performance of processes in terms of efficiency and robustness, as in features matching.

The research of the so-called tie points, useful to obtain the relative orientation of images, as well as the dense cloud reconstruction, are based on specific algorithms able to determine the correspondence among images, taking into account the feature matching or the area based matching techniques, respectively. These operations are strictly related to the scene understanding and interaction, as well as the Simultaneous Localization And Mapping (SLAM; Alsadik and Karam, 2021; Stachniss et al., 2016) technique, i.e., a process in which

a robot localizes itself in an unknown environment while building a map of this environment in real time without any a-priori information.

These methods are described in the following (2.1), together with the homography (2.3), an image processing technique existing in parallel in both Photogrammetry and CV fields, and other procedures (2.2) applied in supporting the tools implementation.

2.1. SCENE UNDERSTANDING

In this section, three topics are presented: feature and area based matching techniques and SLAM. The feature based matching technique is intended to automatically extract features, well recognizable on the frames, compare them among different images and determine the correspondences. This approach is used in selecting the tie points to reciprocally orient images, during the so-called *alignment* phase, i.e., the aerotriangulation obtained by BBA. Tie points are *corners* well recognizable since they represent a discontinuity in the image in both longitudinal and transversal directions and they own peculiar radiometric properties (intensity or RGB – Red Green Blue). The output in photogrammetric software packages consists in a sparse cloud of the 3D positions of these points. The described method (Carrasco et al., 2021; Shao et al., 2021; Liu et al., 2018) could be implemented following different approaches, e.g., the Harris detector (2.1.1) and the SIFT (Scale Invariant Feature Transform - 2.1.2).

On the contrary, area based matching algorithms (Re et al., 2014; Qiu and Zhang, 2015) allow for searching the homologous of a point on a frame considering the radiometric values of its neighbourhood. Thus, it is possible to generate a dense point cloud, matching the points belonging to the considered region. There are local (2.1.3), global and semi-global (2.1.4) methods.

Finally, the principles of SLAM technique, a fundamental approach in mobile robotics applications aiming to obtain a globally consistent estimation of the trajectory and map, are described in Section 2.1.5.

2.1.1. Harris detector

In the Harris detector (Harris and Stephens, 1988), the use of a grayscale 2-dimensional image identified by I is assumed. An image patch pertaining to a window and defined as $(x, y) \in W$ is considered, while another one results from the shift of the quantity $(\Delta x, \Delta y)$. The sum of squared differences (SSD), denoted as *f*, between the initial patch and the shifted one is given by:

$$f(\Delta x, \Delta y) = \sum_{(x_k, y_k) \in W} \left(I(x_k, y_k) - I(x_k + \Delta x, y_k + \Delta y)\right)^2 \quad (2.1)$$

For each generic (x_k, y_k) belonging to the window W. The term $I(x_k + \Delta x, y_k + \Delta y)$ could be approximated by a Taylor expansion. I_x and I_y denote the partial derivatives of I, such that:

$$I(x+\Delta x,y+\Delta y)\approx I(x,y)+I_x(x,y)\Delta x+I_y(x,y)\Delta y \tag{2.2}$$

It produce:

$$f(\Delta x, \Delta y)\approx \sum_{(x_k,y_k)\in W} (I_x(x,y)\Delta x-I_y(x,y)\Delta y)^2 \tag{2.3}$$

The same relation could be expressed by using the matrix M:

$$f(\Delta x, \Delta y)\approx(\Delta x\ \Delta y)M \begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix} \tag{2.4}$$

M represents the structure tensor and it is a second moment matrix obtained by image derivatives.

$$M=\sum_{(x_k,y_k)\in W}\begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2\end{bmatrix} \tag{2.5}$$

Considering an horizontal section of $f(\Delta x, \Delta y)$, it is represented by an ellipse. The resulting semi-axes of the ellipse allow the interpretation of the matrix M in the region of interest:

- *flat region*: both eigenvalues of the matrix are small;
- *edge region*: a large eigenvalue is associated with the gradient at the edge, while in the other direction the eigenvalue is small.
- *corner region*: both eigenvalues will be large.

2.1.2. SIFT

In the SIFT approach (Lowe, 2004), mainly used in SFM, the features are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint. The cost of extracting features is minimized by taking a cascade filtering approach, in which the more expensive operations are applied only at locations that pass an initial test. The main steps needed to generate the set of image features are:

- *Scale-space extrema detection*: searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation.
- *Keypoint localization*: at each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected according to the measures of their stability.
- *Orientation assignment*: one or more orientations are assigned to each keypoint location taking into account the local image gradient directions. All the further operations are performed on image data relatively transformed to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.
- *Keypoint descriptor*: the local image gradients are measured at the selected scale in the region around each keypoint. The best candidate match for each keypoint is found by identifying its nearest neighbour in the database of keypoints from training images. The nearest neighbour is defined as the keypoint with minimum Euclidean distance for the invariant descriptor vector.

Concerning the mathematical formulation of this algorithm, the first stage of keypoint detection is the identification of locations invariant to scale change of the image by searching for stable features across the scale spaces, which are defined as a function, $L(x, y, \sigma)$, obtained from the convolution of a variable-scale Gaussian, $G(x, y, \sigma)$, with an input image, $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x,y) \tag{2.6}$$

where $*$ is the convolution operation in x and y, and:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{2.7}$$

The efficient detection of stable keypoint locations in scale space (Lowe, 1999) is achieved by means of scale-space extrema in the difference-of-Gaussian function convolved with the image, $D(x, y, \sigma)$, which can be computed from the difference of two nearby scales separated by a constant multiplicative factor k:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{2.8}$$

Once a keypoint candidate has been found by checking the local maxima and minima of $D(x, y, \sigma)$, the next step is to perform a detailed fit to the nearby data for location, scale, and ratio of principal curvatures, in order to reject such points having low contrast or poorly localized along an edge. A Taylor expansion is used up to the quadratic terms of the scale-

space function, $D(x, y, \sigma)$, shifted so that the origin is at the sample point (Brown and Lowe, 2002).

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \tag{2.9}$$

where D and its derivatives are evaluated at the sample point and $x = (x, y, \sigma)^T$ is the offset from this point. The location of the extremum, $\hat{x}$, is determined by taking the derivative of this function with respect to x and setting it to zero:

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \tag{2.10}$$

If the offset $\hat{x}$ is larger than 0.5 in any dimension, the extremum lies closer to a different sample point; thus, the sample point is changed and the interpolation is performed for the new point. The final offset $\hat{x}$ is added to the location of its sample point to get the interpolated estimate for the location of the extremum.

The following step is the elimination of the strong edge responses of the difference-of-Gaussian function. A poorly defined peak in the difference-of-Gaussian function will have a large principal curvature across the edge, while a small one in the perpendicular direction. The principal curvatures can be computed from a 2x2 Hessian matrix, H, obtained at the location and scale of the keypoint as:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{2.11}$$

The derivatives are estimated by taking differences of neighboring sample points. The eigenvalues α and β of H are proportional to the principal curvatures of D. Their sum is obtained from the trace of H and their product from the determinant:

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta \tag{2.12}$$

$$Det(H) = D_{xx} D_{yy} - (D_{xy})^2 = \alpha\beta \tag{2.13}$$

In case the determinant is negative, the curvatures have different signs, thus the point is discarded as not being an extremum. Let r be the ratio between the largest magnitude

eigenvalue and the smaller one, so that $\alpha = r\beta$, both Tr(H) and Det(H) depend only on the ratio of the eigenvalues rather than their individual values.

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \tag{2.14}$$

The quantity $(r + 1)^2/r$ is at a minimum when the two eigenvalues are equal and it increases with r. Therefore, using the following relation it is possible to check if the ratio of principal curvatures is below some superimposed threshold, r.

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r} \tag{2.15}$$

Thanks to the assignment of a consistent orientation to each keypoint based on local image properties, the keypoint descriptor can be represented relative to this orientation acquiring invariance to image rotation. The scale of the keypoint is used to select the Gaussian smoothed image, L, with the closest scale, thus all computations are performed in a scale-invariant way. For each image sample, $L(x, y)$, at this scale, the gradient magnitude, $m(x, y)$, and orientation, $\theta(x, y)$, are obtained as:

$$m(x,y) = \sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2} \tag{2.16}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}\right) \tag{2.17}$$

An orientation histogram is formed from the gradient orientations of sample points within a region around the keypoint. Each value of the histogram is weighted by its gradient magnitude and by a Gaussian-weighted circular window with a σ that is 1.5 times than the scale of the keypoint. Peaks in the orientation histogram correspond to dominant directions of local gradients. The highest peak in the histogram is detected, together with any other local peak within 80% of the highest peak, used to create a keypoint with that orientation. Therefore, for locations with multiple peaks of similar magnitude, there will be multiple keypoints created at the same location and scale but different orientations.

Finally, a descriptor is computed for the local image region that is highly distinctive, yet is as invariant as possible to remaining variations, e.g., illumination or 3D viewpoint. Trilinear interpolation is used to distribute the value of each gradient sample into adjacent histogram bins, in order to avoid all boundary effects in which the descriptor abruptly changes because a sample shifts smoothly from being within one histogram to another or from one orientation to

another. Thus, each entry into a bin is multiplied by a weight of $(1 - d)$ for each dimension, where d is the distance of the sample from the central value of the bin. The obtained descriptor consists in a vector containing the values of all the orientation histogram entries. Finally, the feature vector is modified to reduce the effects of illumination change by normalizing it to unit length. In facts, a change in image contrast in which each pixel value is multiplied by a constant will multiply gradients by the same constant, thus the contrast change is eliminated by vector normalization. A brightness change in which a constant is added to each image pixel doesn't affect the gradient values, since they are computed from pixel differences. Therefore, the descriptor is invariant to affine changes in illumination. However, in case of non-linear illumination changes, which cause a large variation in relative magnitudes for some gradients affecting less the gradient orientations, the influence could be reduced by thresholding the values in the unit feature vector to be maximum 0.2, and then renormalizing to unit length. In this way, the distribution of orientations has greater emphasis.

2.1.3. Local methods

Concerning the area based methods, the reference image is called master, while the secondary one is the slave. The examined region in the master image is the template; it represents the neighbourhood of the selected point. The analysed portion of the slave image is called patch and it corresponds to the region of interest whereby calculating the likelihood with respect to the template. If the similarity of the patch is high, it could be supposed that the central points of the corresponding regions are homologous.

The likelihood between the patch and the template could be evaluated by means of different coefficients, e.g., SSD (Sum of Squared Difference) and SAD (Sum of Absolute Difference), not suitable when there are changes in lighting and contrast, or NCC (Normalized Cross-Correlation) and least squares matching, typically used in SFM.

$$SSD = \sum_i \sum_j (f(j,i) - g(j+\Delta x, i+\Delta y))^2 \tag{2.18}$$

$$SAD = \sum_i \sum_j |f(j,i) - g(j+\Delta x, i+\Delta y)| \tag{2.19}$$

$$NCC = \frac{\sum_i \sum_j (f(j,i) - \bar{f}) \cdot (g(j+\Delta x, i+\Delta y) - \bar{g})}{\left(\sum_i \sum_j (f(j,i) - \bar{f})^2 \cdot \sum_i \sum_j (g(j+\Delta x, i+\Delta y) - \bar{g})^2 \right)^{1/2}} \tag{2.20}$$

where f and g are the grey tones of the template and of the patch, respectively, while $\bar{f}$ and $\bar{g}$ are the average of the grey tones of the template and of the patch, respectively.

When the patch is subject to both radiometric (r_0, r_1) and geometric transformations (a_0, a_1, a_2) and (b_0, b_1, b_2) with respect to the template (further than a simple translation manageable with NCC), the least squares approach is the best method to use.

$$f(j,i)+v=r_0+r_1 g\big(j(u,v),i(u,v)\big)=\bar{g}(r_0,r_1,a_0,a_1,a_2,b_0,b_1,b_2) \tag{2.21}$$

Local methods are sensitive to occlusions, which determine changes in the corresponding regions when there are objects standing at different levels of depth.

This gap could be bypassed using global methods.

2.1.4. Global and semi-global methods

Global methods overcome the limits of local ones in the estimation of discontinuities and occlusions in images. All the image pixels are connected by means of a global cost function.

Global algorithms allow to obtain more accurate results, although a higher computational cost is required. In order to approach the issue with the same method but with a lower complexity, semi-global method is intended to decrease the problem dimensionality using a sub-system of image points called scan line.

In the semi-global approach, the global cost C is decreased along one of the dimensional paths, which are joint in each pixel and pass through the image. Analysing a scan line along a certain direction, the matching cost of a point p for each value of shift d is evaluated, taking into account also the cost at the previous step $(p-1)$ and the related disparity d'.

$$C(p,d)=C_{match}(p,d)+\min\begin{cases} C(p-1,d) \\ C\big(p-1,d'\big)+P,\ d'\neq d \end{cases} \tag{2.22}$$

Thus, the first term $C_{match}(p,d)$ concerns the local likelihood, while the second considers a correlation among neighbours. The coefficient P represent a fixed penalty to pay whenever there is a change in depth along the minimum path, in order to remove surges due to noise or ambiguities, favouring the surfaces with constant disparity.

In Table 2-1, a resume about the comparison of local, global and semi-global method is shown.

Table 2-1. Features of local, global and semi-global methods.

	Features
Local methods	- Use of delimited correlation windows to identify the correspondences among homologous points, independently of each other - Fast and suitable for real-time applications - Point maps not dense because of uniform regions and occlusions
Global methods	- Assignment of disparity such that the global cost function extended to all pixels of the images is minimized - Dense maps of points - High computational cost
Semi-global methods	- Minimization of a semi-global cost function to reduce the problem complexity - Bond between accuracy and speed

2.1.5. SLAM

As already stated, SLAM (Durrant-Whyte et al., 1996; Durrant-Whyte and Bailey, 2006) is a method which goal is the simultaneous estimation of both the object position and the mapping of the surrounding environment without any a-priori information. Considering a robot moving in a scene while collecting relative observations of several unknown landmarks by using an embedded sensor, at a time instant k the following quantities are defined: x_k, the state vector describing the location and orientation of the vehicle; u_k, the control vector, applied at time $k-1$ to drive the vehicle to a state x_k at time k; m_i, a vector describing the location of the i^{th} landmark whose true location is assumed time invariant; z_{ik}, an observation taken from the vehicle of the location of the i^{th} landmark at time k. Furthermore, also the following sets, including the previously specified quantities for each instant k, are defined: $X_{0:k}$, the history of vehicle locations; $U_{0:k}$, the history of control inputs; m, the set of all landmarks; $Z_{0:k}$, the set of all landmark observations.

SLAM algorithm consist of two recursive steps, i.e., the sequence of prediction (called time-update and described as probability distribution in equation (2.23)) and correction (known as measurements-update and described in equation (2.24)).

$$P(x_k,m|Z_{0:k-1},U_{0:k},x_0)=\int P(x_k|x_{k-1},u_k)\times P(x_{k-1},m|Z_{0:k-1},U_{0:k-1},x_0)dx_{k-1} \tag{2.23}$$

$$P(x_k,m|Z_{0:k},U_{0:k},x_0)=\frac{P(z_k|x_k,m)P(x_k,m|Z_{0:k-1},U_{0:k},x_0)}{P(z_k|Z_{0:k-1},U_{0:k})} \tag{2.24}$$

Equations (2.23) and (2.24) provide a recursive procedure for the calculation of the joint posterior $P(x_k, m|Z_{0:k}, U_{0:k}, x_0)$ related to the robot state x_k and map m at a time k based on all observations $Z_{0:k}$ and all control inputs $U_{0:k}$ up to and including time k. The recursion is a function of a vehicle model $P(x_k|x_{k-1}, u_k)$ and an observation model $P(z_k|x_k, m)$.

Several solutions to the probabilistic SLAM problem are possible, aiming to find an appropriate representation for both the observation model and the motion model that allows an efficient and consistent computation of the prior and posterior distributions in (2.23) and (2.24). One of the most common methods is the extended Kalman filter (EKF), suitable to treat a state-space model with additive Gaussian noise. The basis for the EKF-SLAM method is to describe the vehicle motion in the form:

$$P(x_k|x_{k-1},u_k) \Leftrightarrow x_k=f(x_{k-1},u_k)+w_k \tag{2.25}$$

In (2.25), the function f models vehicle kinematics, while w_k represent additive, zero mean uncorrelated Gaussian motion disturbances with covariance Q_k. Conversely, the observation model is described as:

$$P(z_k|x_k,m) \Leftrightarrow z_k=h(x_k,m)+v_k \tag{2.26}$$

In (2.26), the function h describes the geometry of the observation, while v_k are additive, zero mean uncorrelated Gaussian observation errors with covariance R_k.

Thus, it is possible to compute the mean (2.31) and the covariance (2.32) of the joint posterior distribution $P(x_k, m|Z_{0:k}, U_{0:k}, x_0)$ from the time-update ((2.27) and (2.28)) and the observation-update ((2.29) and (2.30)) terms:

$$\hat{x}_{k|k-1}=f(\hat{x}_{k-1|k-1},u_k) \tag{2.27}$$

$$P_{xx,k|k-1}=\nabla f P_{xx,k-1|k-1}\nabla f^T+Q_k \tag{2.28}$$

$$\begin{bmatrix}\hat{x}_{k|k}\\ \hat{m}_k\end{bmatrix}=[\hat{x}_{k|k-1} \quad \hat{m}_{k-1}]+W_k[z_k-h(\hat{x}_{k|k-1},\hat{m}_{k-1})] \tag{2.29}$$

$$P_{k|k}=P_{k|k-1}-W_k S_k W_k^T \tag{2.30}$$

obtaining, respectively:

$$\begin{bmatrix}\hat{x}_{k|k}\\ \hat{m}_k\end{bmatrix}=E\left[\begin{matrix}x_k\\ m\end{matrix}|Z_{0:k}\right] \tag{2.31}$$

$$P_{k|k}=\begin{bmatrix}P_{xx} & P_{xm}\\ P_{xm}^T & P_{mm}\end{bmatrix}_{k|k}=E\left[\begin{pmatrix}x_k-\hat{x}_k\\ m-\hat{m}_k\end{pmatrix}\begin{pmatrix}x_k-\hat{x}_k\\ m-\hat{m}_k\end{pmatrix}^T|Z_{0:k}\right] \tag{2.32}$$

In (2.28), ∇f is the Jacobian of f evaluated at the estimated $\hat{x}_{k-1|k-1}$. In (2.30), S_k and W_k can be written as:

$$S_k=\nabla h P_{k|k-1}\nabla h^T+R_k \tag{2.33}$$

$$W_k=P_{k|k-1}\nabla h^T S_k^{-1} \tag{2.34}$$

where ∇h is the Jacobian of h evaluated at $\hat{x}_{k|k-1}$ and $\hat{m}_{k-1}$.

The presented EKF-SLAM method is commonly used for solutions applied to navigation or tracking issues, although it has some criticalities concerning the computational effort, growing quadratically with the number of landmarks, the sensitivity to incorrect association of landmarks observations, and the non-linearity.

2.2. IMAGE PROCESSING TECHNIQUES

Image processing techniques include pre-processing and filtering, edge detection, segmentation, mathematical morphology, texture analysis, model matching, contextual and multi-scale techniques.

Among the others, the edge detection, the segmentation, the erosion and dilation morphological operators, and the Hough transform data model comparison techniques are deepen since they have been used in the implementation of the proposed tools.

2.2.1. Edge detection

Edge detection allow the recognition of the object boundary by means of the identification of discontinuities in intensity.

One strategy is to compute the image gradient $\nabla I(x,y)$ by means of the image partial derivatives I_x and I_y:

$$\nabla I(x,y)=\left[\frac{\partial I}{\partial x},\frac{\partial I}{\partial y}\right]^T=\left[I(x,y)_x,I(x,y)_y\right]^T=\left[I_x,I_y\right]^T \tag{2.35}$$

The gradient points towards the direction θ of the highest change in intensity, obtainable by means of the following relation:

$$\theta=\tan^{-1}\left(\frac{\partial I}{\partial x}\Big/\frac{\partial I}{\partial y}\right) \tag{2.36}$$

The local contrast $\|\nabla I\|$ along the edge normal, i.e., the vector pointing in direction of the maximum intensity change, is defined as magnitude:

$$\|\nabla I\|=\sqrt{\left(\frac{\partial I}{\partial x}\right)^2+\left(\frac{\partial I}{\partial y}\right)^2} \tag{2.37}$$

Thanks to these parameters, it is possible to easily determine the presence of edges by setting a threshold, such that if the pixel magnitude is higher than this value an edge point is recognized.

Noise constitutes a substantial problem in edge detection, in facts, edge detectors are conceived in order to strongly respond to sharp changes and, in parallel, noise values are typically uncorrelated to the neighbours ones, so that they are very different. In order to avoid this issue, two convolutions are used, the former is applied to smooth the image, while the latter to differentiate.

An example of this strategy in edge detection is represented by the Sobel-Feldman (Duda & Hart, 1973) edge detector, resulting by the multiplication of a 1D Gaussian filter by the x (2.38) or y (2.39) derivative:

$$\begin{bmatrix}1\\2\\1\end{bmatrix}\begin{bmatrix}-1 & 0 & 1\end{bmatrix}=\begin{bmatrix}-1 & 0 & +1\\-2 & 0 & +2\\-1 & 0 & +1\end{bmatrix} \tag{2.38}$$

$$\begin{bmatrix}-1\\0\\1\end{bmatrix}\begin{bmatrix}1 & 2 & 1\end{bmatrix}=\begin{bmatrix}-1 & -2 & -1\\0 & 0 & 0\\+1 & +2 & +1\end{bmatrix} \tag{2.39}$$

Other strategies contemplate the use of the second derivative, so that zero-crossing mark the edge location, like in the Marr and Hildreth (Marr & Hildreth, 1980) approach.

The edge detector employed in the implementation of the tools discussed in this thesis is the Canny's one (Canny, 1986). It consists of three steps: the application of the derivative of Gaussian along the x and y axes to filter the image; the computation of the image gradient; the calculation of the magnitude and the orientation of the gradient. A thinning operation is performed in order to suppress the pixels in the gradient magnitude map that do not represent local maximum (non-maxima suppression). In other words, it is checked if each pixel is a local maximum in its neighbourhood in the direction of gradient or not and, if it represent a local maximum, it is retained as an edge pixel, otherwise suppressed.

Two threshold are established, so that the values above the highest limit are always kept, conversely the ones below the lower limit are always discarded. The remaining values, within the range of the two thresholds, are considered valid only if they belong to a path that join confirmed edges with values above the highest limit, so that a connected contour is linked.

2.2.2. Segmentation

The main objective of segmentation (Nikhil & Sankar, 1993) is to use image features to map single pixels into homogeneous groups of pixels called regions. The set of regions that derives from the segmentation process is called partition. The recognition problem consists in five steps: conditioning, labelling, grouping, extracting, and matching.

Concerning conditioning, the image could be modified to remove the information that is not of interest or to enhance certain images features, by means of removing noise filtering, increasing contrast, and reconstructing missing parts. The information contained in the image could be considered as a composition of spatial events, which corresponds to sets of connected pixels. The operation of labelling determines for each pixel the event in which it is involved. A typical example of labelling on grey level images is determined by the threshold application, which assigns a label value equal to 1 to all pixels with a larger grey level compared to the threshold value. The results is represented by a binary image, in which the value 1 indicates pixels that belong to the object of interest whereas all the other pixels are assigned label 0.

After the labelling operation, which tags pixels as pertaining or not to an object, the operation of grouping allows to join pixels with the same label in a cluster. This operation generates regions, i.e., a set of neighbouring pixels, even called connected components. They are created using a suitable operator, requiring a binary image as input and returning a symbolic image as output. The label assigned to each pixel is an integer that identifies the belonging region of the pixel.

Two pixels p and q with the same label f belongs to the same connected component C if there is a sequence of points $(p_0, p_1, \dots, p_n)$ of value f belonging to C where $p_0 = p$ and $p_n = q$ and p_i is neighbor to p_{i-1} for $i = 1, \dots, n$.

Depending on the employed operator, there are 4-connected or 8-connected regions.

The iterative algorithm suitable to perform the grouping consists of three steps:

- A label is assigned to each pixel with value 1.
- The value of each pixel is replaced by the smallest label value (different from zero) of its neighbours; the pixel above and the one on the left are considered, obtaining a 4-connected grouping.
- Step two is carried out recursively from the left to the right and from the top to the bottom. Thus, bottom-up row scan follows, considering the 4-connected neighbourhood made by the lower pixel and the right one.

The second and third steps are iterated. The procedure ends when there are no more label changes.

Another method, called classical algorithm, is based on similar principles. Like in the iterative algorithm, a 4-connected neighbourhood is examined by considering the pixel above and the one on the left of the analysed pixel. Each pixel value is replaced by the label value (different from zero) of its neighbours. When two different labels could be assigned to one pixel, the lower label value is chosen and declared equivalent to the other, listing them in a specific table. Thus, equivalence classes are defined; a label value corresponding to the lowest label is assigned to each class.

Further operations on built regions could be performed, such as extracting and matching. For each region identified through grouping, various properties could be extracted, e.g., area, spatial moments, etc.; these features depend on the grey levels of the pixels belonging to the region and their spatial arrangement.

A meaning can be given to a region using a perceptual grouping, which allows to interpret the set of pixels as an already known object. Matching operation determines the interpretation of a set of pixels, associating 3D objects or 2D shapes.

2.2.3. Erosion and dilation

Erosion (Figure 2-1) and dilation (Figure 2-2) (Soille, 2004) are basic morphological operators applied to binary images. The processing concerns the shape of the objects and it is based on the definition of a structuring element B_x centered on the point x.

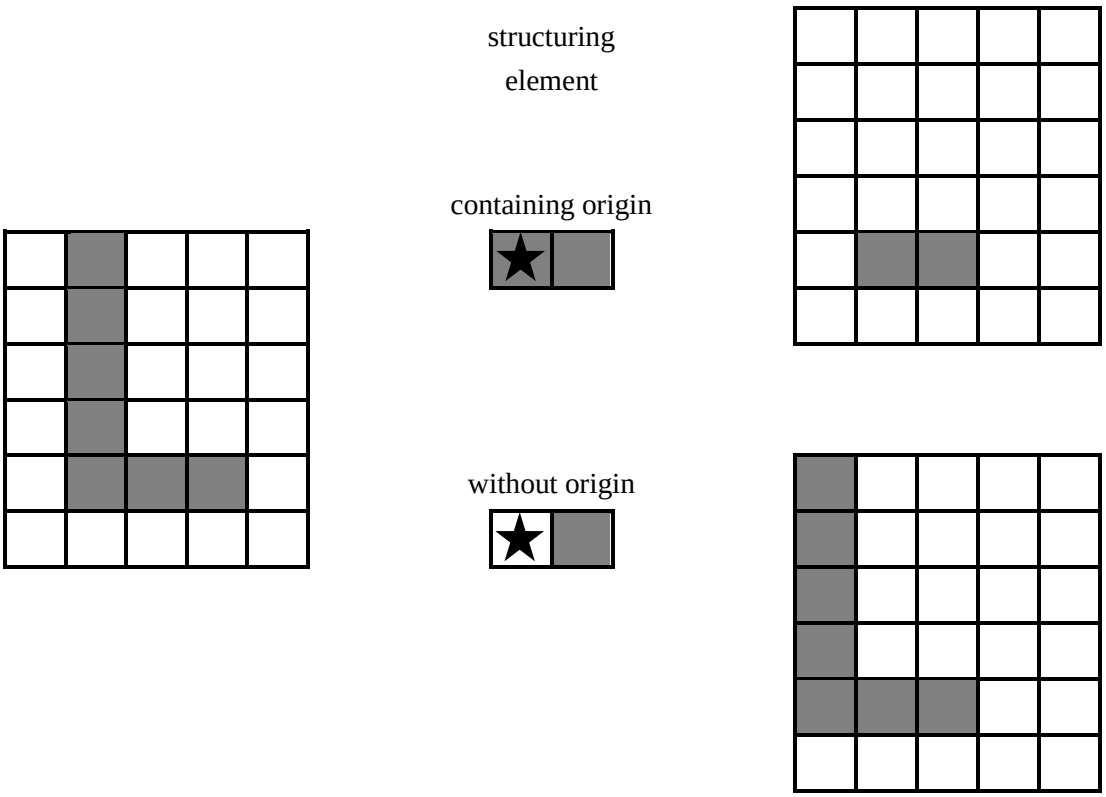


Figure 2-1. Binary erosion.

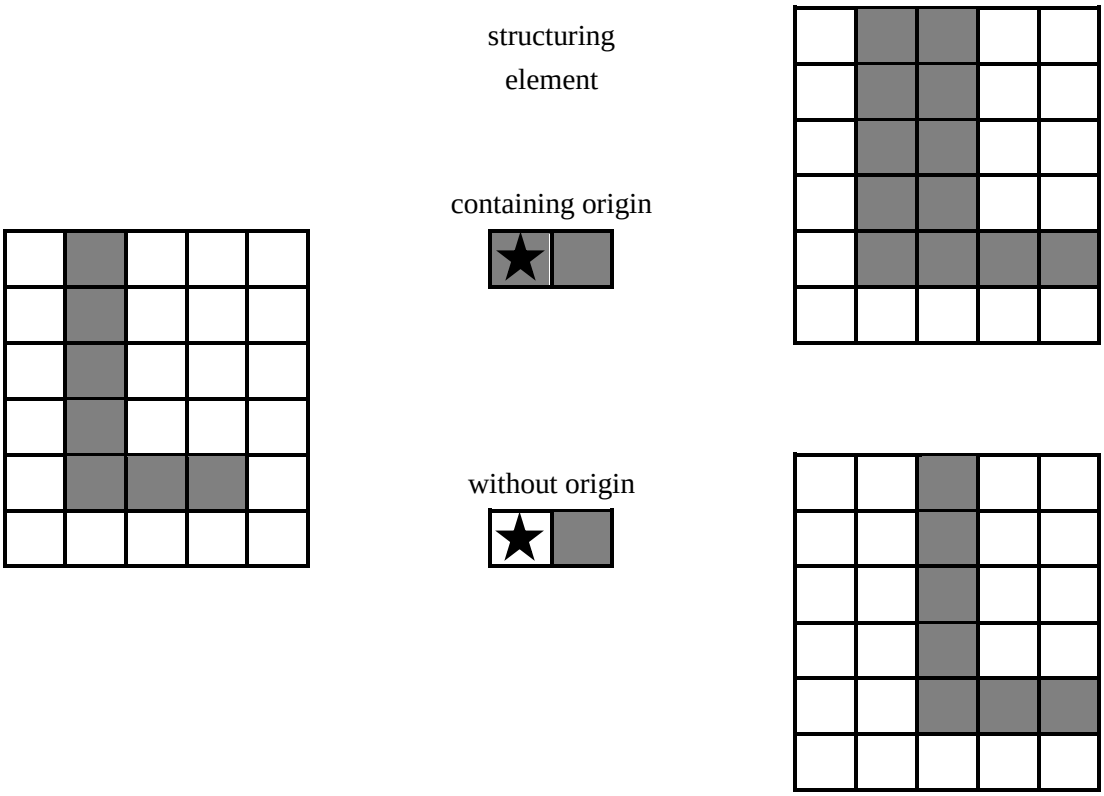


Figure 2-2. Binary dilation.

The erosion (2.40) is typically used to remove pixels from the boundaries of objects in the image. For each pixel of A, centre the structuring element $\overleftrightarrow{B_x}$; set the current pixel to full (1) if all full pixel of $\overleftrightarrow{B_x}$ correspond to full in the analysed portion of A, otherwise set it to empty (0).

The dilation (2.41) is typically used to add pixels to the boundaries of objects in the image. For each pixel of A, centre the transposed structuring element $\overleftrightarrow{B_x}$; set the current pixel to full (1) if at least one full pixel of $\overleftrightarrow{B_x}$ correspond to full in the analysed portion of A, otherwise set it to empty (0).

Erosion and dilation are invariant regarding translation. They are not inverse of each other.

$$A \ominus B = \left\{ x \middle| \overleftrightarrow{B_x} \subseteq A \neq 0 \right\} \tag{2.40}$$

$$A \oplus B = \left\{ x \middle| \overleftrightarrow{B_x} \cap A \neq 0 \right\} \tag{2.41}$$

The properties of erosion and dilation are resumed in the following Table 2-2.

Table 2-2. Erosion and dilation properties.

#	Condition	Property
1	$B \supset B'$	$(A \ominus B) \subset (A \ominus B')$ $(A \oplus B) \supset (A \oplus B')$
2	$0 \notin B$	$(A \ominus B) \subset A \subset (A \oplus B)$
3	$A_1 \supseteq A_2$	$(A_1 \oplus B) \supseteq (A_2 \oplus B)$ $(A_1 \ominus B) \supseteq (A_2 \ominus B)$
4	$\forall A,B$	$(A \oplus B) = (B \oplus A)$
5	$\forall A,B,C$	$(A \oplus B) \oplus C = A \oplus (B \oplus C)$ $(A \ominus B) \ominus C = A \ominus (B \ominus C)$

The succession of the two operators could lead to the opening or closing operations: the former is obtained by one erosion and one dilation and allows to remove small details and to disjoint weakly linked objects; the latter is obtained by one dilation and one erosion and permits the union of close or weakly connected objects. Both opening and closing operations are idempotent, i.e., they can be applied multiple times without changing the result beyond the initial application.

2.2.4. Hough transform

The Hough transform (Hough, 1962) is a feature extraction technique used to detect curves in images, converting coordinates from Cartesian to polar. Any point within the image space is represented by a sinusoidal curve in the Hough space. In addition, two points in a line segment generate two curves, which are overlaid at a location corresponding to a line through the image space.

Considering the problem for straight lines localization in an image, the Hough Space is a 2D plane, which horizontal and vertical axis represent, respectively, the slope and the intercept of a line represented in the form $y = ax + b$. One line on the analysed image produces a point on the Hough Space, since a line is characterized by its slope a and intercept b. On the other hand, a point (x_i, y_i) on the image can have an infinite number of lines pass through it. Therefore, a point produces a line in the Hough Space in the form of $b = ax_i + y_i$ (Leavers, 1992). In the Hough Transform algorithm, the Hough Space is used to determine whether a line exists in the analysed image.

2.3. HOMOGRAPHY

A homography (Ondrašovič and Tarábek, 2021; Kahl et al., 2008), or projective transformation, is a type of transformation that maps straight lines onto another plane. It could

be applied in image rectification, image registration, or computation of camera motion (rotation and translation) between two images.

The homography, represented by a non-singular 3×3 matrix *H*, relates the pixel co-ordinates of two images.

$$x' = H x \tag{2.42}$$

When this transformation is applied to each pixel of the input image, the result is a warped version corresponding to the second provided frame. Two images are related by a homography if and only if both images are viewing the same plane from a different angle and they are taken from the same camera, rotated around its centre of projection without any translation. Note that the homography relationship is independent of the scene structure.

Among their properties, it is worth noting that the inverse and the composition of homography matrices are homographies themselves.

H is a homogeneous transformation with eight degrees of freedom (dof). It is a composition of transformations, including in hierarchical order isometries, similarities, affinities, and projectivities. These transformations are described in Table 2-3.

Table 2-3. Transformation scheme.

	Features
Isometries	- transformations of the euclidean plane preserving the euclidean distance $\begin{pmatrix} x'_1 \\ x'_2 \\ 1 \end{pmatrix} = \begin{bmatrix} \epsilon\, cos\theta & -sin\theta & t_x \\ \epsilon\, sin\theta & cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}, \epsilon = \pm 1$ $H_E = \begin{bmatrix} R & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$ - 3 dof → 2 points correspondences needed for the computation - invariants: length, angles, areas
Similarities	- isometries + isotropic scaling $\begin{pmatrix} x'_1 \\ x'_2 \\ 1 \end{pmatrix} = \begin{bmatrix} s\, cos\theta & -s\, sin\theta & t_x \\ s\, sin\theta & s\, cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}$ $H_S = \begin{bmatrix} sR & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$ - 4 dof → 2 points correspondences needed for the computation - invariants: angles, ratios of length, ratios of areas
Affinities	- non-singular linear transformation followed by a translation $\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ $H_A = \begin{bmatrix} A & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$ - 6 dof → 3 points correspondences needed for the computation - invariants: parallel lines, ratios of length of parallel line segments, ratios of areas
Projectivities	- a general non-singular transformation of homogeneous coordinates $H_P = \begin{bmatrix} A & \boldsymbol{t} \\ \mathbf{v}^T & v \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ v_1 x_1 + v_2 x_2 \end{pmatrix}$ - invariants: cross-ratio

An affinity could always be decomposed as rotations and non-isotropic scaling.

$$A=R(\theta)R(-\varphi)DR(\varphi) \tag{2.43}$$

This follows from the SVD (Singular Value Decomposition) of *A*.

$$A=UDV^{T}=(UV^{T})(VDV^{T}) \tag{2.44}$$

In the estimation of homographies, the minimal problem takes into account four points correspondences. A popular algorithm used to obtain the *H* matrix is the Direct Linear Transformation (DLT; Fangi, 1995; El-Ashmawy, 2018). The relation (2.45) contains the lines of the DLT matrix for each point correspondence.

$$\begin{bmatrix} 0 & 0 & 0 & -w_i'x_i & -w_i'y_i & -w_i'z_i & y_i'x_i & y_i'y_i & y_i'z_i \\ w_i'x_i & w_i'y_i & w_i'z_i & 0 & 0 & 0 & -x_i'x_i & -x_i'y_i & -x_i'z_i \end{bmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \end{pmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{2.45}$$

$$H=\begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \tag{2.46}$$

A robust estimation could be achieved by means of the RANSAC method (RANdom SAmpling Consensus; Fischler and Bolles, 1981), an iterative model able to identify and exclude outliers.

Homography is used in both Photogrammetry and Computer Vision. In particular, in the former case DLT algorithm is applied to camera calibration, in order to obtain the IO parameters as a function of the *h* coefficients, and for the achievement the so-called aerotriangulation. In the latter scenario, homography is at the basis of visual odometry, a technique employed in robotics to estimate the motion of a camera in real time using sequential images.

A SUITE IN SUPPORT OF THE PHOTOGRAMMETRIC WORKFLOW

The focus of this thesis work is about a suite of tools designed to support the photogrammetric workflow during all its phases.

Three procedures, presented in Table 3-1, have been conceived in order to introduce significant novelties in the consolidated approach of photogrammetric software, from the planning of the survey to the outputs restitution.

Table 3-1. Reference photogrammetric workflow steps and related tools.

Steps	Tools
planning	→ U.Ph.O. (Unmanned Photogrammetric Office)
acquisition	→ BBA real-time procedure
post-processing	
restitution	→ MAGO (Adaptive Mesh for Orthophoto Generation)

Particular attention has been paid to the final precision and to the management of complex scenarios.

Concerning U.Ph.O. (Passoni et al., 2018), it allows the planning of the survey basing on a network simulation approach, which takes into account the real visibility and occlusions; the final output is an estimation of the expected precision of the survey, whose level of detail is set according to the resolution of the input DSM.

The BBA real-time procedure has been conceived to exploit the Bundle Block Adjustment principles in real-time navigation, positioning and modelling. It is intended to work in indoor environments, where the support of GNSS is denied. Two branches are being developed, both for real-time purposes: the former aims to compute the positioning of the shooting cameras (EO), whose RO is fixed and known and which are joint with a moving vehicle, whether car or robot or UAV; the latter seeks to update the 3D model of a moving object acquired by stationary sensors.

Finally, MAGO is an innovative tool whose approach is based on the use of the point cloud, instead of the mesh, as input source for the orthophoto reconstruction (Gagliolo et al., 2019a and 2019b). The algorithm has been particularly conceived to treat with a high level of automation the production of orthophotos pertaining vertical façades. This aim has pushed the author to enrich the U.Ph.O. procedure adding a specific module to deal with scenarios characterised by a vertical development, in order to obtain the estimation of precision in such portions of particular interest.

The detailed description of each tool is reported in the following paragraphs. In particular, the illustration of the software features and principles is organised putting as first MAGO, in order to have an overall point of view starting from the final objective of the photogrammetric workflow. Then, a mention to the ongoing work about the BBA real-time procedure is reported, focusing on the so-called “active” part of the strategy dedicated to navigation and positioning. Finally, the

improvement of U.Ph.O. with the addition of a module for the estimation of precision in such scenarios with vertical development is presented.

3.1. MAGO NOVELTY IN ORTHOPHOTOS GENERATION APPROACH

Orthophotos represent a typical output of a photogrammetric survey and, thanks to the diffusion of specific software packages, their generation and usage have become even more widespread. They represent a suitable instrument to perform high-precision measurements, thanks to the uniform scale given by the orthogonal projection. For this reason, they are widely employed in several fields, mainly in cartography, but also in environmental and building engineering, cultural heritage, precision farming and forest management.

Nevertheless, there are some issues about the accuracy of orthophoto reconstruction, since it is often downgraded by the introduction of meshes and DSMs to be used as surfaces representing the object. The use of a more accurate and reliable input, such as a point cloud, makes these approximations avoidable. Furthermore, the procedure has been specifically enhanced in order to automatically create “unrolled” joint orthophotos in presence of adjacent walls.

The input data of the procedure are the initial user-defined orthophoto plane, the image and its IO and EO parameters, together with a point cloud representing the object. Each pixel of the image is projected on the orthophoto plane at its original resolution via an iterative process, which builds an adaptive mesh, defined by means of the three best fitting points, where the collinearity rays and the point cloud intersect.

3.1.1. State of the art in orthophoto production

Thanks to the development of innovative techniques and technologies in the photogrammetric field, orthophotos quality is increasing. The problem of reliability in orthophotos production has given rise to the research about True Orthophotos© (Amhar, 1998; Habib et al., 2007; Wang et al., 2018; Li et al., 2020), which are focused on the combination of two masked orthophotos that represent buildings and terrain of an urban area separately, to avoid misrepresentation and lack of information in hidden areas. Based on the described criterion, some tools have been implemented to generate accurate orthophotos, e.g. TORPEDO (Three dimensional Object Resource Package for Enhancing Digital Orthophotos; Amhar, 1998), ACCORTHO (ACCurate ORTHOprojection; Boccardo et al., 2001), GCOrtho (Geometrically Corrected Orthophotos; Barazzetti et al., 2007).

Many photogrammetric software packages, both commercial (e.g., Agisoft Metashape©, ContextCapture™, Pix4D©, LiMapper©, Ortro©) and open source (MicMac, OpenDroneMap) include a specific step for orthophotos reconstruction, typically at the end of their workflow, when the polygonal mesh is already available (Sai et al., 2019; Ngadiman et al., 2018).

Further than these two methods, orthophotos could be obtained even using front views of the point cloud, according to a defined point of view, or realizing them with the traditional technique, starting from a previously computed DSM.

In these last approaches, the highest resolution is in the order of the point cloud mean spacing; using photogrammetric software packages, the highest resolution is comparable with the Ground Sample Distance (GSD) size, but the object description is depleted due to a simplified approach to realize the polygons of the mesh.

In this context, MAGO has been developed to avoid this rough simplification by means of a step-by-step adaptive mesh directly obtained from the point cloud. Moreover, it allows the orthophoto reconstruction using the highest possible resolution, in the same order of the GSD of the original image.

3.1.2. MAGO approach and workflow

The MAGO approach for orthophoto generation is designed to consider a specific triangular plane area where the image pixel is projected at its original resolution, avoiding the approximation and loss of definition typically introduced by the mesh reconstruction phase.

The MAGO workflow consists in the following phases:

- definition of the orthophoto plane;
- acquisition of IO and EO parameters and images themselves from external sources;
- definition of the orthophoto dimensions and resolution;
- iterative process to determine the three best-fitting points, that define the plane where the collinearity ray and the point cloud intersect;
- projection of the corresponding colour of each pixel in the image on the orthophoto plane.

Firstly, the point cloud representing the object, produced by external photogrammetric software processing or acquired by TLS or LiDAR, has to be uploaded in MAGO.

The IO and EO parameters of the images, previously obtained using an external software, are given as input, together with the images themselves.

The user must define the initial orthophoto plane, via the coordinates of three points expressed in the point cloud RS in case of a generic plane, or via the default options in case of planes parallel to XY, XZ or YZ. The point cloud is roto-translated in a service RS according to the set orthophoto plane. The geometric transformations applied allow obtaining a new triplet of Cartesian axes: XY axes define the chosen plane and Z results as orthogonal direction, according to the right-handed convention. The new so-obtained RS permit the easy determination of which point in each cell is visible from the orthophoto plane, taking into account the new Z coordinate.

The orthophoto resolution is user-defined, and it should be adequately chosen considering the GSD of the image, in order to get an optimal final result.

The orthophoto dimensions could be set, if a limited portion of the input point cloud needs to be considered, or could be automatically computed by MAGO considering the minimum and maximum values along each dimension. In this case, the highest possible orthophoto dimensions are:

$$\begin{aligned} \frac{X_{MAX} - X_{MIN}}{\text{resolution}} &= \text{orthophoto width} \\ \frac{Y_{MAX} - Y_{MIN}}{\text{resolution}} &= \text{orthophoto height} \end{aligned} \tag{3.1}$$

Then, a regular grid is created to organize the input points; the details about the criteria at the basis of the spatial resolution computation, automatically done by the software, are reported in the following paragraph 3.1.3.

Finally, the user could set a downscaling factor for the original image in order to speed up the procedure and to match the orthophoto resolution.

All these settings can be inserted via a Graphical User Interface (GUI), as depicted in Figure 3-1.

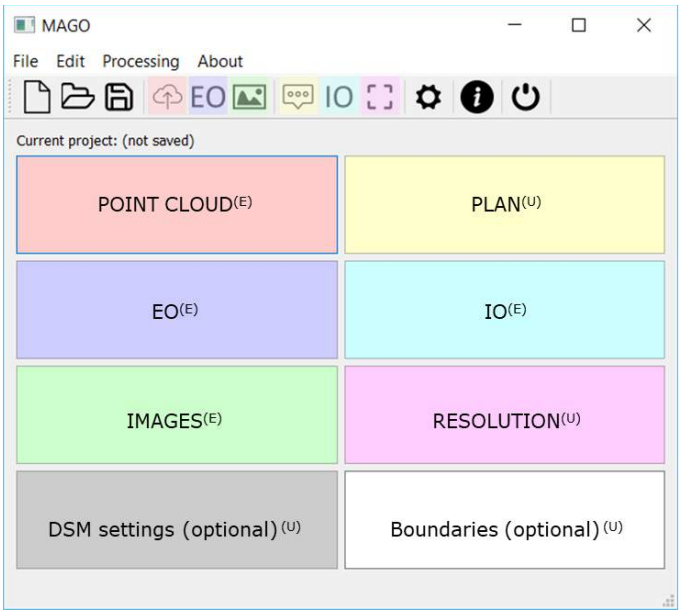


Figure 3-1. MAGO's Graphical User Interface.

Once the input parameters have been defined, the orthophoto is produced by means of an iterative process to determine the plane where the collinearity ray from each pixel and the point cloud intersect. This plane is one of the faces of the adaptive mesh obtained by the original point cloud. Each face of the adaptive mesh is generated directly from the input point cloud, without any further simplification, re-sampling or approximation of the point cloud itself, thus, the adaptive mesh is the highest-resolution mesh possible.

To build the orthophoto, the procedure starts from the collinearity equations (1.1) applied on each pixel on the image.

At the beginning of the iterative process, the first attempt Z_P is imposed equal to Z_{MAX}, which represents the maximum height of the points in the point cloud, having preferably excluded and filtered out the outliers. X_P and Y_P are consequently computed, applying the collinearity equations.

The so-obtained $\underline{X}_P$ coordinates match a cell of the grid, defined in the preliminary phase of the procedure, and its corresponding pixel on the orthophoto plane.

Considering the candidate cell of the grid, the correspondence is confirmed if there is a point falling inside the volume defined by the cell planimetric dimensions and a threshold along z, termed δz and fixed on the basis of the point cloud spacing s.

In this case, two possible scenarios are considered, according to the distance between the found point and the collinearity ray. If the distance is lower than a threshold given by the GSD size affected by the downscaling factor, the point itself is considered as a correspondence between the image and the orthophoto pixel, and it is projected on the orthophoto plane. Conversely, in case the threshold is exceeded, the algorithm searches in the cell itself and in its eight neighbours two points to build a triangle with the first one.

The chosen criteria to define the best-fitting triangle are delineated in the following. Due to the point cloud inhomogeneity, the points that are excessively near or far from the first found one are excluded to avoid an unrealistic description of the surface.

A threshold is defined as a ring with internal (r_i) and external (r_e) radius respectively of 40% and 180% of the grid cell size.

To avoid the definition of a sub-vertical plane, a cylinder is introduced as follows: the base is the previously defined ring, while the height (h) varies within a distance proportional to the external diameter of the ring and the tangent of ±80°. Moreover, at least one of the two candidate points has to be on the opposite side of the track of collinearity ray on the orthophoto plane, with respect to the first one. Finally, the found vertices have to respect a minimum projected area of the triangle equal to 20% of the regular grid cell area. Thus, the intersection between the adaptive triangle, defined by the three best fitting points, and the collinearity ray determines the final correspondence between the image pixel and the orthophoto one.

The chosen criteria and thresholds are depicted in Figure 3-2.

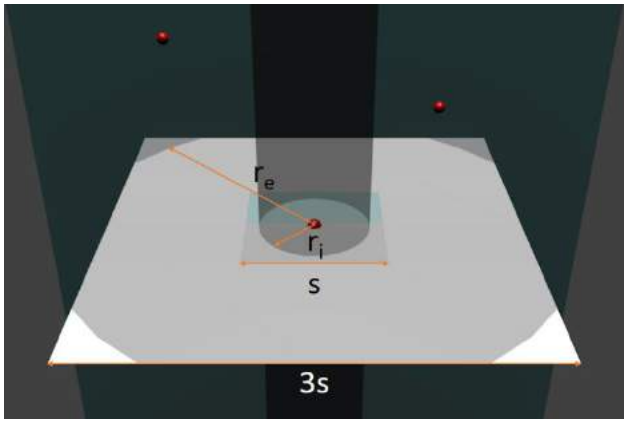


Figure 3-2 Criteria and parameters of adaptive meshing and matching iterative process.

In case the threshold δz is not complied or the cell is empty, Z_P is automatically decreased of the spacing s and the updated Z_P value is used for the iterative process, until a correspondence is found. If the value of Z_{MIN} (representing the minimum height of the points in the point cloud) is reached without having found a matching, the pixel is discarded and the procedure continues with the analysis of a new one, until all the pixels on the image have been analysed.

Figure 3-3 depicts the scheme of MAGO workflow.

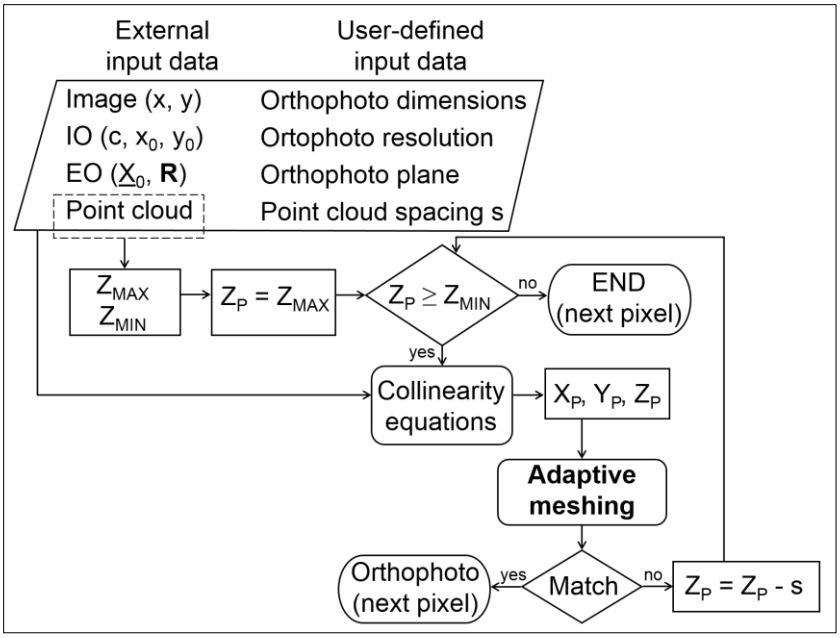


Figure 3-3. MAGO workflow.

The novelty of the approach at the basis of the proposed method is the use of the point cloud, instead of the mesh, as input source for the orthophoto reconstruction. This condition makes possible to consider each time a specific triangular plane area resulting from the three best-fitting points of the cloud, without any superimposed a priori choice. Thus, the image pixel is projected at its original resolution, avoiding the approximation and loss of definition typically introduced by the building mesh phase.

The so-conceived tool has been recently enhanced introducing new functionalities and improving the computational effort. At the state of the art, the processing time is comparable with the consolidated software packages; the memory management has been made more effective, even thanks to the use of the open source library OpenCV (OpenCV Development Team, 2019) and to the implementation of a function for the automatic organization of the point cloud in a spatial grid. Moreover, an innovative module for the automatic creation of "unrolled" joint orthophotos of adjacent walls has been developed.

3.1.3. Evolution of implemented versions

As already stated, the implementation of MAGO has been fulfilled in C++ environment; the code structure includes about 3500 lines of code, together with a simple Graphical User Interface (GUI) realized in Qt.

An initial approach considered the employment of some existing open source tools: Cimg Library (2019), Image Magick (2019), and matrix.h (2019). Cimg Library and Image Magick were used to operate images, also in different file formats, while matrix.h allowed an ease management and calculation of data stored in matrices.

During its development, the tool has been significantly improved, thanks to an optimized computational approach and to new functionalities. In particular, the code has been fully revolutionized introducing the open source library OpenCV, which stands for Open Computer

Vision. The contribution of this novelty has been very significant in the optimization because OpenCV works with both 2D and 3D entities and it is capable of managing both matrices and images using the same variable *cv::Mat*, allowing a simpler code architecture.

Moreover, three new functions have been introduced: the first is needed to apply a service roto-translation in a new local Reference System (RS), in order to easily permit the recognition of the visible portions from any orthophoto plane. The second is intended to manage more effectively the input point cloud, organizing it in a spatial grid according to optimized parameters automatically based on its content. Finally, the third is needed to set the chosen plane in parallel to the analysed façade time by time.

Furthermore, the realization of an orthomosaic starting from multiple input images, averaging the values from the source pixels, has been sketched.

Two preliminary steps are required ahead of the iterative phase for the determination of the three best-fitting points; they are represented by the new functions for the change of the RS according to the orientation of the orthophoto plane (paragraphs 3.1.5 and 3.1.6) and for the automatic organization of the point cloud in an optimized spatial grid (3.1.4). These operations help in the selection of the points visible from the orthophoto plane, in order to use them for the adaptive mesh generation.

3.1.4. Creation of the supporting spatial grid

The spatial grid, in which the point cloud is stored, is created testing iteratively the cell dimension until the ideal one is found; the criterion taken into account in the determination of the best condition is the rate of empty cells, excluding opportunely the borders and the internal regions larger than a certain size, which are probably effectively empty.

To determine the cell side (s), the procedure starts imposing it equal to the ratio between the rectangular area (A), limited by the planimetric boundaries, and the number of points (n_P).

$$s = \frac{A}{n_P} \tag{3.2}$$

If the points are regularly distributed, this value generates a grid containing cells with a point numerousness equal to 1 per each and without unoccupied places. In practice, the value is accepted as ideal if the empty cells are below the 5% of the total, after deduction of the borders and the large internal regions, distinguished using an image processing segmentation technique called *grouping*; otherwise, an iterative analysis is started.

The *grouping* segmentation technique, presented in the paragraph 2.2.2, is intended to create regions starting from pixels containing similar or equal values, according to a pre-set threshold.

This procedure has been applied considering as input image the grid containing the cells numerousness. In particular, since the main interest is related to the isolation of the empty cells, it has been applied considering zero and non-zero values.

Once the regions are composed, all the pixels connected to the zero values on the borders are counted as out of the point cloud boundaries. In this way, the real boundaries generated by the point cloud 3D shape are considered.

The internal zero-regions are identified as empty regions if their member-cells are more than 20; this quantity is absolute, unlike the other parameters assumed in the processing that are relative, since already the referring space grid size is determined ad hoc for the case study. The empty cells not belonging to the borders and to the zero-regions larger than 20 members are considered as valid empty cells.

At the first attempt, a new spacing value is calculated: the total area of the cells decreased by the valid empty ones is divided by the number of points of the cloud. If the difference between this new value and the previous is less than the 10% of the initial spacing, the second attempt is made considering the calculated value. If this condition is not satisfied, a new test is introduced to check if the valid empty cells are below the 5% of the total. In the following iterations, only the latter test is maintained. If the condition is not satisfied, the spacing value is increased of its 25%.

All the mentioned tolerance values, including both percentages and other parameters, have been determined empirically with the aim of generalizing the approach, optimizing it on the basis of the treated point cloud.

It is worth noting that the first value, already described as the ideal one in presence of regularly distributed points, since it would generate a grid containing cells with a point numerousness equal to 1 per each and without unoccupied places, is also the lowest possible. Thus, the changes operated with the iterations are only incremental.

3.1.5. Geometric transformations to deal with generic orthophoto planes

The relationship between the internal RS of the image x^* and the external one of the object X has been expressed in the expression (1.3). Since a new external RS is used to treat the point cloud, considering the XY plane parallel to the orthophoto plane, the object coordinates X_P and the camera shooting centre X_0 need to be expressed using this new convention; thus, a rotation matrix is applied, as showed in (3.3).

$$\begin{bmatrix} X_{new} \\ Y_{new} \\ Z_{new} \end{bmatrix} = R_{ROT} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3.3}$$

Consequently, the equation (1.3) is transformed as follows in order to keep the correct relationship with the image coordinates:

$$\begin{bmatrix} x^* \\ y^* \\ z^* \end{bmatrix} = \lambda R_C \cdot R_{ROT}^{-1} \cdot R_{ROT} \begin{bmatrix} X_P - X_0 \\ Y_P - Y_0 \\ Z_P - Z_0 \end{bmatrix} \tag{3.4}$$

where the product $R_{ROT}^{-1} \cdot R_{ROT}$ represent the identity matrix. Thus, the new rotation matrix that relates the new object system with the image system is obtained as:

$$R_{Cnew}=R_C \cdot R_{ROT}^{-1} \tag{3.5}$$

How is possible to obtain the proper R_{ROT} matrix for the analyzed orthophoto plane?

The input plane is identified using three points, in particular: the first is the origin of the new RS; the second belongs to the new X axis; the third belongs to the new Y axis. Since no transformations have been applied yet, their coordinates need to be provided in the original external RS.

Starting from these points, the new X and Y axes will be identified by vectors obtained as:

$$\overline{X_{new}}=\frac{(X_{pt2}-X_{pt1},\ Y_{pt2}-Y_{pt1},\ Z_{pt2}-Z_{pt1})}{\sqrt{(X_{pt2}-X_{pt1})^2+(Y_{pt2}-Y_{pt1})^2+(Z_{pt2}-Z_{pt1})^2}}$$

$$\overline{Y_{new}}=\frac{(X_{pt3}-X_{pt1},\ Y_{pt3}-Y_{pt1},\ Z_{pt3}-Z_{pt1})}{\sqrt{(X_{pt3}-X_{pt1})^2+(Y_{pt3}-Y_{pt1})^2+(Z_{pt3}-Z_{pt1})^2}} \tag{3.6}$$

Z axis is consequently identified as the normal to the just established plane, according to the right-handed convention: the vector components are equal to the plane coefficients a, b and c.

$$ax+by+cz+d=0 \tag{3.7}$$

$$\overline{Z_{new}}=\frac{(a,b,c)}{\sqrt{a^2+b^2+c^2}} \tag{3.8}$$

In order to obtain from the three provided points the implicit form of the plane, it is sufficient to impose the following condition:

$$\det\begin{pmatrix} x-X_{pt1} & y-Y_{pt1} & z-Z_{pt1} \\ X_{pt2}-X_{pt1} & Y_{pt2}-Y_{pt1} & Z_{pt2}-Z_{pt1} \\ X_{pt3}-X_{pt1} & Y_{pt3}-Y_{pt1} & Z_{pt3}-Z_{pt1} \end{pmatrix}=0$$

$$\begin{aligned} a&=(Y_{pt2}-Y_{pt1})*(Z_{pt3}-Z_{pt1})-(Z_{pt2}-Z_{pt1})*(Y_{pt3}-Y_{pt1}) \\ b&=(Z_{pt2}-Z_{pt1})*(X_{pt3}-X_{pt1})-(X_{pt2}-X_{pt1})*(Z_{pt3}-Z_{pt1}) \\ c&=(X_{pt2}-X_{pt1})*(Y_{pt3}-Y_{pt1})-(Y_{pt2}-Y_{pt1})*(X_{pt3}-X_{pt1}) \\ d&=-aX_{pt1}-bY_{pt1}-cZ_{pt1} \end{aligned} \tag{3.9}$$

It could be noticed in equations (3.6) and (3.8) that the vectors need to be normalized, i.e., to be divided by their module to obtain norm equal to 1, not least in order to maintain the correct scale for the point cloud transformation.

Once these operations are done, it is possible to write down R_{ROT} as:

$$R_{ROT} = \begin{bmatrix} \overline{X_{new}} \\ \overline{Y_{new}} \\ \overline{Z_{new}} \end{bmatrix} \tag{3.10}$$

i.e., the lines of the matrix are represented by the normalized vectors.

R_{ROT} is made by a set of vectors representing an orthonormal basis, since they are all unit vectors and orthogonal to each other. For this reason, it is an orthogonal matrix, which own the property of equivalence of the inverse R_{ROT}^{-1} and the transpose R_{ROT}^{T}.

Furthermore, it is a proper rotation matrix since its determinant is equal to 1.

From the equation (3.5) is possible to infer the new rotation angles with respect to the new axes. The Cardano rotation matrix is built considering that the object is fixed, the reference frame is rotating, and the object is described in the rotated reference frame. The three concatenated elementary rotations are applied as reported in the convention of equation (1.4), obtaining the *R* previously detailed in equation (1.8). The equation (1.9) is applied to R_{Cnew} matrix to obtain the new ω, φ and κ angles for the transformed RS.

Once the RS has been transformed in the new service one, the point cloud is organized in a spatial grid, according to the aforementioned criteria. Whenever multiple rotations are applied in order to adapt the plane in parallel to the analysed façade or to turn into the following side (3.1.6), the spatial grid size (3.1.4) is maintained equal to the first estimation, in order to treat homogeneously the whole case study.

3.1.6. Automatic check and adaptation of the orthophoto plane in parallel to the analysed façade

The creation of an "unrolled" joint orthophoto of adjacent walls is supported, beyond the already described functions, even by a dedicated module designed to evaluate the parallelism between the current orthophoto plane and the analysed façade. The check is based on the assessment of the Z coordinate homogeneity: a histogram is plotted, in order to support the establishment of the specific operative condition.

The first scenario concerns the situation with a single peak, which confirms the real homogeneity in the Z coordinate, corresponding to the parallelism between the analysed plane and the wall. Thus, the current plane is definitively applied as orthophoto plane. By means of an operative sequence consisting in combinations of dilation and erosion, i.e., closing in mathematical morphology (2.2.3; *cv::dilate* and *cv::erode* functions in OpenCV library), applied to obtain sharper borders, then Canny edge detection (2.2.1; *cv::Canny*) and, consequently, of Hough Transform (2.2.4; *cv:: HoughLinesP*), the boundaries of the area with homogeneous Z

coordinate are computed and identified as discontinuities. Along these lines, the orthophoto plane will be rotated to face the adjacent walls.

The second scenario contemplates the presence of multiple peaks; it represents a condition in which the plane is parallel to the façade, but the analysed surface is not regular because of differences in depth. Thus, the procedure focuses step by step on the discontinuities per each "rung", analysing if the variation is symptomatic of the development of a hidden surface normal to another direction, which links the two levels of depth.

The third scenario represent the unchecked parallelism. The resulting histogram does not present any peak; thus, it is possible to infer that the Z coordinate is not homogeneous. The angle between the current plane and the façade direction is computed in order to repeat the procedure putting the plane precisely in front of the analysed wall.

3.1.7. Mention to the use of MAGO for semantic segmentation

Taking inspiration from the module for automatic adaptation of the orthophoto plane in parallel to the analysed façade, MAGO has been further enriched of a draft for an extra module aiming to reach semantic segmentation in urban scenarios (Gagliolo and Sguerso, 2021).

Semantic segmentation is a Computer Vision technique that aims to the recognition and the comprehension of the content of an image at the pixel level. This approach is widely used in remote sensing applications, especially in the analysis of urban scenarios (Ajmar et al., 2019; Huang et al., 2019, Schmitz et al., 2019, Zhou et al., 2019) or in the delineation of forest trees (Chen et al., 2021; Sothe et al., 2020; Kempf et al., 2019).

The segmentation approach could be based on imagery (Marmanis et al., 2018) or three-dimensional models (Ao et al., 2019), as well as on the combination of both 2D and 3D information (Ding et al., 2019). Typically, deep learning methods are applied to such procedure, including, to cite some examples, Conditional Random Fields (CRF; Pan et al., 2020; Lafferty et al., 2001), Markov Random Fields (MRF; Zoltan and Josiane, 2012), Spatial Pyramid Pooling (SPP; Zhengyu and Joohee, 2020), and Convolutional Neural Networks (Cresson, 2020; Martinez-Soltero et al., 2020; Ouyang and Li, 2021).

The proposed approach is based on a machine-learning algorithm, which combines on the one hand 2D and 3D information, while on the other hand geometric and radiometric criteria, in order to obtain a unique classification.

Starting from the 3D point cloud given as input for MAGO procedure, two raster maps are produced: the former consists in the Digital Surface Model (DSM) of the scene, while the latter represents the corresponding nadir greyscale map. Both these raster images contain Not a Number (NaN) values where the cell could not be filled with any source information from the 3D point cloud.

Both the raster maps are processed using the Harris Corner detector (2.1.1), by means of the corresponding function implemented in OpenCV. This technique allows to rate each image pixel with a mark *R*, according to the presence of a large variation in intensity with respect to the neighbours. In particular, the pixels are associated to the following groups basing on the obtained mark *R*:

- $R > 0$: corner, i.e., significant change in all directions;

- $R < 0$: edge, i.e., no change along the edge direction;
- $|R|$ small: flat region, i.e., no change in all directions.

The value R is obtained from the second moment matrix M (2.5), which is computed from image derivatives:

$$M=\begin{bmatrix} \sum_{x,y} W(x,y)\cdot I_x^2(x,y) & \sum_{x,y} W(x,y)\cdot I_x(x,y)I_y(x,y) \\ \sum_{x,y} W(x,y)\cdot I_x(x,y)I_y(x,y) & \sum_{x,y} W(x,y)\cdot I_y^2(x,y) \end{bmatrix} \tag{3.11}$$

$$M=\begin{bmatrix} \sum I_xI_x & \sum I_xI_y \\ \sum I_xI_y & \sum I_yI_y \end{bmatrix} = \sum \begin{bmatrix} I_x \\ I_y \end{bmatrix} \cdot [I_x \quad I_y] = \sum \nabla I \cdot (\nabla I)^T \tag{3.12}$$

Each horizontal section of the SSD function $f(\Delta x, \Delta y)$ (2.4), which could be expressed in terms of M, represent the equation of an ellipse. The diagonalisation of the M matrix allows to obtain the lengths of the ellipse axes and their orientation, by means of the eigenvalues λ_1 and λ_2 and the corner response measure R, respectively.

$$M=R^{-1}\cdot\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\cdot R \tag{3.13}$$

The method implemented in OpenCV takes into account the corner response measure, using the value R, which is calculated as:

$$R=\lambda_1\lambda_2-\alpha(\lambda_1+\lambda_2)^2=\det(M)-\alpha\,\mathrm{trace}(M)^2 \tag{3.14}$$

where α is an empirically determined constant ranging within 0.04 and 0.06. In the present case, the value 0.04 has been adopted.

The input parameters for the *cv::cornerHarris* function are the source image, the destination image, the kernel size, the aperture parameter for the Sobel operator, and the constant α.

Once both the geometric and the radiometric input maps have been processed with this technique, each cell has been classified according to the obtained R value. In particular, the service images containing the R marks obtained from the radiometric and the geometric contribution are called R_{colour} and R_{geom}, respectively. Moreover, a synthesis of the two contributions is stored in the matrix R_{class}, giving a label based on the following criterion, such that the obtainable R_{class} values are resumed in Table 3-2.

Table 3-2 R_{class} attribution conditions.

if			**then**
$R_{colour} > 0$	&	$R_{geom} > 0$	$R_{class} = +4$
$R_{colour} < 0$	&	$R_{geom} < 0$	$R_{class} = -4$
$R_{colour} < 0$	&	$R_{geom} > 0$	$R_{class} = +3$
$R_{colour} > 0$	&	$R_{geom} < 0$	$R_{class} = -3$
$\|R_{colour}\|$ small	&	$R_{geom} > 0$	$R_{class} = +2$
$R_{colour} > 0$	&	$\|R_{geom}\|$ small	$R_{class} = -2$
$\|R_{colour}\|$ small	&	$R_{geom} < 0$	$R_{class} = +1$
$R_{colour} < 0$	&	$\|R_{geom}\|$ small	$R_{class} = -1$
$\|R_{colour}\|$ small	&	$\|R_{geom}\|$ small	$R_{class} = 0$

Resuming, the labelling operation allows identifying the level of variability associated to each pixel, considering the changes in X and Y directions and in geometric and radiometric information.

The following step is the processing of the classification map by grouping in homogeneous regions the neighbour pixels with the same assigned value, by means of the technique described in paragraph 2.2.2. The set of regions that derives from this segmentation process is called partition.

While in paragraph 3.1.4 the segmentation regarded the distinction of empty and full cells, in the present case the labelling results from the *R*-based classification determine for each pixel the so-called event in which it is involved, tagging if pixels are pertaining or not to an area. Then, the operation of grouping allows joining the pixels with the same label in a cluster.

Until now, the greyscale colour space has been chosen to apply the geometric segmentation using the Harris corner detector for the gathering of the discontinuities. Nevertheless, this colour space does not easily allow recognizing the hue of the analysed pixel, as well as the well-known RGB (Red Green Blue). Thus, it has been decided to convert in the HSV (Hue Saturation Value) range the original map of colours, obtained from the coloured point cloud. Starting from the HSV associated to each cell, several masks are arranged in order to identify the pixel membership. In particular, the threshold criteria are listed in the following; they have been chosen according to a personal interpretation and not using a superimposed classification. The OpenCV interpretation of the input values is due to the bytes coverage and requires that the H is within 0° and 180° instead of 360°, and S and V are within 0 and 255 instead of between 0 and 1. Table 3-3 resumes the input parameters for colour masking, according to OpenCV convention.

Table 3-3. HSV categories.

Colour	H	S	V
White	0 – 180	0 – 24	230 – 255
Red	0 – 14	25 – 255	100 – 255
	165 – 180	25 – 255	100 – 255
	0 – 14	100 – 255	25 – 255
	165 – 180	100 – 255	25 – 255
Brown	0 – 14	25 – 99	25 – 99
Purple	135 – 180	25 – 99	25 – 99
Green	45 – 74	25 – 255	25 – 255
	15 – 44	25 – 99	25 – 99
Blue	105 – 134	25 – 255	25 – 255
Cyan	75 – 104	25 – 255	25 – 255
Yellow	15 – 44	25 – 255	100 – 255
	15 – 44	100 – 255	25 – 255
Magenta	135 – 164	25 – 255	100 – 255
	135 – 164	100 – 255	25 – 255
Black	0 – 180	0 – 255	0 – 24
Grey	0 – 180	0 – 24	25 – 230

In such regions having a resulting value R_{class} equal to 0 or 1, i.e., with no significant radiometric and null or low geometric variation respectively, the HSV masking is applied homogeneously, on the basis of the most recurrent value in the area.

In the first step of the segmentation, the criteria coming from R_{class} or directly the height information resulting from the DSM are combined with the colour inferred from the HSV masking, as listed in Table 3-4.

Table 3-4. Applied criteria for the preliminary labelling.

Class	Label	Conditions		
0	no data	NaN		
1	buildings	HSV ∉ green	&	$Z > 6$ m
2	roads	grey	&	$R_{class} < 2$
4	trees	green	&	$R_{class} \geq 2$
5	low vegetation	green	&	$R_{class} < 2$
8	background clutter	remaining		

In this phase, the static scene is distinguished in five macro-areas, including buildings, roads, trees and low vegetation, as well as the remaining background.

Further criteria need to be implemented in order to point out also the three remaining categories, according to the UAVid (Lyu et al., 2020) benchmark classification, i.e., static cars, humans and moving cars, isolating them from the generic background.

In this regard, the actual potentialities of the algorithm are not suitable to discern humans from cars. Thus, the label static cars (class 3) and moving cars (class 7) are changed to static and moving objects respectively, while the category humans (class 6) is suppressed.

Regarding the segmentation of static objects, they are extracted from the generic background checking the presence of at least a certain number of cells labelled as road (class 2) or static object

(class 3) in the neighbourhood of the analysed pixel by using a kernel. In particular, the road and the static object cells need to be more than the half of cells filled with categories different from the background and the empty ones (classes 8 and 0, respectively).

If at least one of the neighbour cells is road, a further check on the difference between the analysed pixel and the mean of the heights in the surrounding road cells is performed, i.e., if the Z coordinate of the analysed pixel is lower than three meters over the road average height, the matching with class 3 is confirmed.

The last step is the recognition of the moving objects, which is achieved thanks to the comparison of images from different epochs. These frames, obtained from the acquisition of the camera as central projections, are orthogonally projected by using the proper module of MAGO, firstly on a plane with a similar orientation to the original image attitude, then on the XY plane. The intermediate phase, which takes into account a service plane approximately at the same inclination of the original image, allows MAGO to optimize the research of the matching points that compose the adaptive mesh.

Once the projections are performed, the resulting greyscale maps are subtracted, in order to highlight the difference from the previous to the following epoch. The pixels that are not visible in both the analysed views are excluded from the comparison. Moreover, a threshold of 30 in the range of greyscale tones is applied to exclude changes barely perceivable by the human eye.

3.2. BUNDLE BLOCK ADJUSTMENT REAL-TIME PROCEDURE

BBA, presented in Section 1.1, is a unified method to estimate simultaneously the IO and EO parameters and the 3D coordinates of the object points in a statistically optimal manner.

The procedure presented in this paragraph, based on BBA approach, has been conceived with a double aim to position and navigate operators indoor and to model the environment, both in real-time. It has been implemented firstly in Matlab® in a preliminary version, then in C++ environment, with the support of existing free and Open Source libraries, i.e., OpenCV, Ceres and librealsense.

The module for navigation and positioning purposes has been implemented within the Finnish Geospatial Research Institute (FGI), where the author spent a period of four months abroad.

In these circumstances, the work was intended to contribute to the project named Real-Time AI-Supported Ore Grade Evaluation for Automated Mining (RAGE), in order to achieve a reliable indoor positioning solution, based on the integration of several technologies, including LIDAR, visual cameras and IMU. The working environment, consisting in a mine, is particularly challenging, since it is completely GNSS denied, dark and prone to fast mutability.

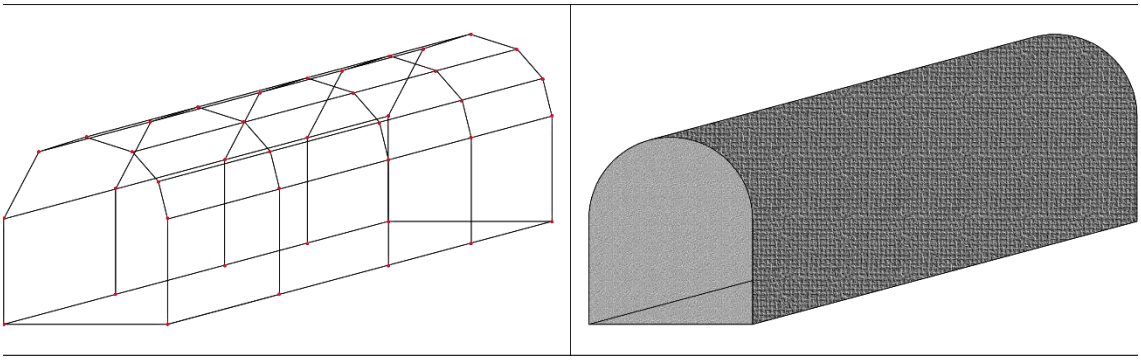


Figure 3-4 On the left, tie points structure needed for navigation; on the right, scheme of the resulting volumetric model of the tunnel.

The issue may be studied from two mirrored perspectives, denominated as active and passive. The former scenario, deepen during the period abroad, aims to obtain the positioning of a moving object; it is more complex, since the sensors themselves move changing their own position in space and time according to the object itself. In the latter situation, the objective is the real-time updated modelling of an evolving scenario; the sensors are fixed and just the surrounding environment is changing instant by instant.

As already stated, the dissertation is focused on mentioning the ongoing work about the navigation module, described in paragraph 3.2.1, while the passive approach represent a further future development of the topics introduced in this thesis.

3.2.1. Navigation and positioning module

The scarcity and the deeper location of mineral deposits, as well as the potential risk exposure for the workers, require the development of increasingly automated techniques for underground mining operations (Lösch et al., 2018).

In order to achieve a precise and reliable positioning in the challenging indoor environment of underground mines, the interest of the researchers is mainly addressed in involving multiple sensors and combining different techniques, such as: sensor network, fingerprinting, Radio-Frequency Identification (RFID), Inertial Measurement Unit (IMU), Very-Low Frequency (VLF), geomagnetic field, Visible Light Communications (VLC), Ultra-Wide Band (UWB), Sound Navigation and Ranging (SONAR), Light Detection and Ranging (LIDAR), and visual cameras (Iturralde et al., 2014; Thrybom et al., 2015; Azhari et al., 2017; Li et al., 2018; Ren et al., 2019).

In this scenario, the RAGE project is intended to continue the work carried out within the FGI during the previous KAIVOS project (Kaasalainen et al. 2018; Malkamäki et al. 2018; Malkamäki et al., 2019). In that case, positioning has been obtained by means of an UWB network (Di Pietra et al., 2018), which is definitely an effective technique for indoor navigation purposes, but requires a huge number of devices installed in the mine in order to reach an acceptable level of accuracy.

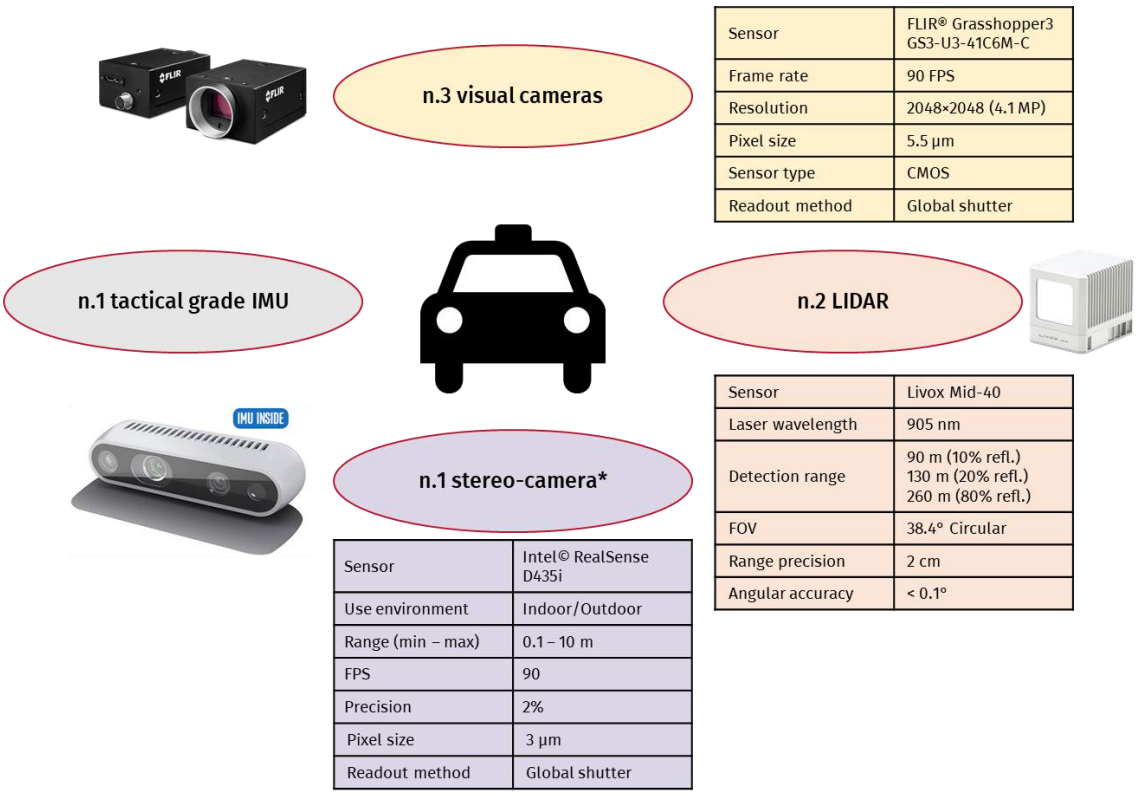
The innovation of the new RAGE project is intended to overcome the necessity of installing expensive infrastructures, which is in conflict with the fast mutability of the mine environment and the risk of collapse for both natural and artificial reasons. In this context, the implementation

of the navigation module has been developed by the author during the period abroad spent at FGI, in order to apply it to indoor positioning in mines.

As already mentioned, in this challenging scenario, the goal of the underway work is to obtain the location of the operator (whether a worker or a robot) by means of the integration of LIDAR, Visual Cameras and IMU techniques. Actually the integration of these sensors is widely studied for navigation purposes in several fields of application (Shao et al. 2019; Zhang and Singh, 2018), including also underground mining. In particular, the MFLO (Map Free LIDAR Odometry) algorithm has been provided by Lesak (2019), in order to extract 3D ego-motion using a range flow constraint equation in real-time. The method is designed to operate in GPS-denied and light-deficient environments; several sensors have been evaluated, including rotating 3D or 2D LIDAR, Time of Flight (TOF) range finder, RGB-D camera, and orientation sensor.

The proposed method for indoor positioning, navigation and 3D modelling is based on the combination of different sensors, listed in the following: two LIDARs Livox, two visual cameras FLIR® Grasshopper3 GS3-U3-41C6M-C, one visual camera FLIR® Flea3 FL3-U3-13E4M-C, one stereo-camera Intel© RealSense D435i, one tactical grade IMU.

The instruments have been mounted on the top of a car by means of appropriate racks, in order to carry out some preliminary tests.



Sensor	FLIR® Grasshopper3 GS3-U3-41C6M-C
Frame rate	90 FPS
Resolution	2048×2048 (4.1 MP)
Pixel size	5.5 µm
Sensor type	CMOS
Readout method	Global shutter

Sensor	Livox Mid-40
Laser wavelength	905 nm
Detection range	90 m (10% refl.) 130 m (20% refl.) 260 m (80% refl.)
FOV	38.4° Circular
Range precision	2 cm
Angular accuracy	< 0.1°

Sensor	Intel© RealSense D435i
Use environment	Indoor/Outdoor
Range (min – max)	0.1 – 10 m
FPS	90
Precision	2%
Pixel size	3 µm
Readout method	Global shutter

Figure 3-5 Selected instrumentation and its features.

The setup has been planned in advance, in order to enhance the effectiveness of the system according to the expected result. Concerning in particular the visual cameras, several different layouts have been studied trying to reach the optimal configuration for the application of both

SLAM (Simultaneous Localization And Mapping) and BBA algorithms. The elements taken into account during the planning are: the maintenance of the needed overlapping among the views, on the one hand by the different cameras involved in the scene reconstruction and on the other hand by the shot sequence; the light conditions; the speed.

In the chosen setup the three cameras are placed on a row at a fixed distance of about 50 cm from each other, pointing at the ground with a proper tilt, in order to avoid to frame the car itself. The so obtained area of interest is illuminated thanks to the car headlights. Moreover, even the alternative to point the cameras to the tunnel sides has been considered, but it would have implicated some difficulties, e.g., the necessity of a higher amount of cameras to connect the two opposite points of view, or the a priori exclusion of one of the two sides, as well as the necessity of additional light sources.

The challenging light conditions pushed the research team to choose monochrome cameras; in facts, grayscale images are more suitable in dark environments and allow a faster processing in real-time conditions.

The maximum speed of the vehicle has been established in light of the set FPS rate and accounting the required percentage of overlapping (80% forward, 60% side); the obtained value is about 90 km/h, which is undoubtedly precautionary in such environment.

The RO and IO of cameras have been estimated in advance, calibrating them with a calibration panel of the proper size.

All the mentioned instruments have to be synchronized to obtain simultaneous acquisitions. The delay in the signal triggering has to be evaluated to understand if it could affect significantly the measurements.

In the described conditions, the aim of the code, based on BBA applied to real-time applications for indoor navigation purposes (Sheta et al., 2012; Jin et al., 2018), is to compute step by step the EO parameters of the cameras. This operation is possible thanks to the knowledge of:

- the IO parameters, obtained during the calibration phase;
- a set of coded targets, used as GCPs and located in easily reachable areas, such as the starting point of the path;
- the matching points among the images (three frames per single instant).

Achieving the EO of the cameras and knowing their relative position with respect to the object of interest, i.e., the car, it is easy to infer its positioning in the scene. Supposing that it is not possible to arrange in advance any target in addition to the ones at the beginning of the path, an increasing drift is expected to affect the measurements. Thus, the GCPs standing in the initial controlled area stiffen the system like in the case of a bracket, using a parallelism with Structural Mechanics, but the longer is the length of the path, the higher is the probability of drift. The schema of such situation is depicted in the left part of Figure 3-6, while the right one is dedicated to the hypothesis of a changed path at a new instant of time T1.

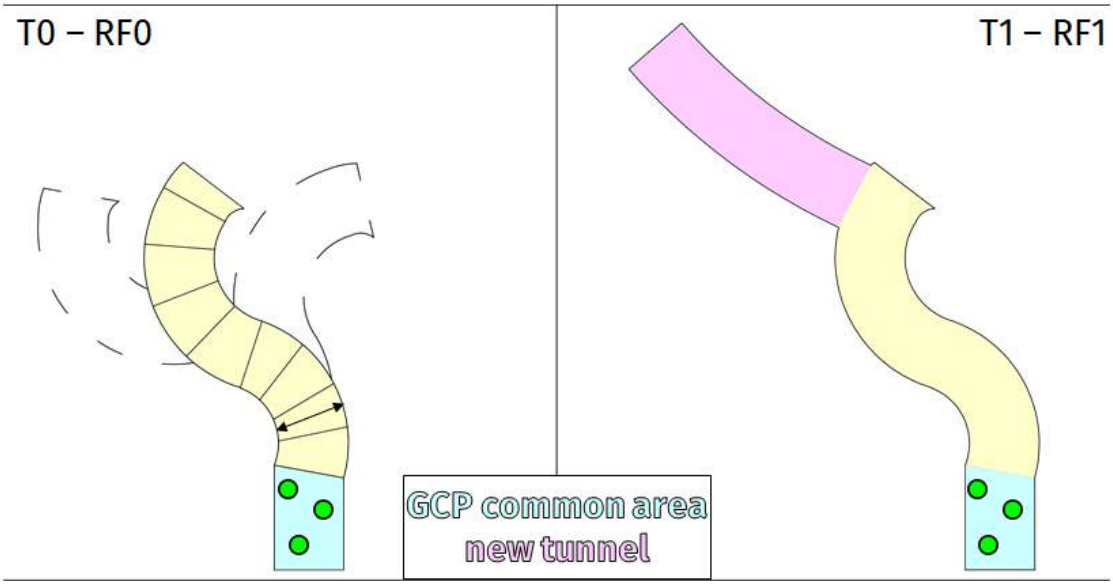


Figure 3-6 Operative scheme in a changing environment.

The scarcity and the location of the GCP, together with the single strip geometry of the photogrammetric block, make necessary to consider the already adjusted tie points as fixed, giving them a weight considerably higher than the others.

The steps of the proposed procedure are basically two, i.e., select the input data by means of a simple Graphical User Interface (GUI) and apply the BBA algorithm on 9 images at time (3 shots per 3 instants).

The required inputs are:

- the IO parameters, obtained by the calibration;
- the GCPs object coordinates and their template for the automatic recognition;
- the list of the images and their estimated EO parameters inferred from the operations in the initialization area (in cyan in Figure 3-6) thanks to the recognition of the GCPs by means of Normalized Cross Correlation (deepen in paragraph 2.1.3).

As already stated, the GCPs are automatically recognized in the images and associated to their corresponding label thanks to the coded shape. The templates are provided in order to apply the Normalized Cross Correlation on the acquired images. The coordinates of the centre of each target are registered and treated like observations, included in the vector *y*. In total, four observation per matched point are available, i.e., ξ and η coordinates in the two matching images.

This formulation takes into account also the distortions and the error associated to the observation.

The unknowns are represented by: the six external orientation parameters of one reference camera, since the relative position and attitude of the others is known in advance with the calibration; the three object coordinates of each matching point.

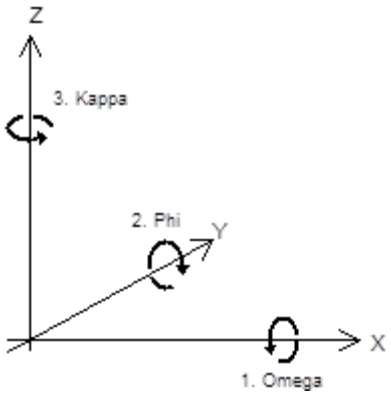


Figure 3-7. Scheme of position and attitude in the reference frame.

The estimated values of the unknowns, contained in the vector *x*, are used as initial guess, then a non-linear least squares procedure is applied iteratively to correct this estimation, as long as the error is higher than an established limit value.

The design matrix, or Jacobean, *J* is *m* rows by *n* columns, corresponding to the number of observations and unknowns, respectively. The difference between *m* and *n* is the redundancy.

The entries of the matrix *J* are calculated using the estimated and the known values and substituting them in the partial derivatives of collinearity equations.

Considering the first group of images, an equal weight is assigned to all the observations, so the matrix P is the identity matrix:

$$N=J^{T}PJ \tag{3.15}$$

Since the J matrix is made by blocks, it is easier and faster to treat them separately, dividing the unknowns of the EO by the ones related to the points.

$$N=\begin{bmatrix}J_1^T\\J_2^T\end{bmatrix}P\begin{bmatrix}J_1 & J_2\end{bmatrix}=\begin{bmatrix}N_{11} & N_{12}\\N_{12}^T & N_{22}\end{bmatrix} \tag{3.16}$$

The steps to obtain the new estimated *x* vector are shown in the following:

$$C=\begin{bmatrix}J_1^T\\J_2^T\end{bmatrix}Py=\begin{bmatrix}C_1\\C_2\end{bmatrix} \tag{3.17}$$

$$\begin{bmatrix}N_{11} & N_{12}\\N_{12}^T & N_{22}\end{bmatrix}\begin{bmatrix}\hat{x}_1\\\hat{x}_2\end{bmatrix}=\begin{bmatrix}C_1\\C_2\end{bmatrix} \tag{3.18}$$

$$\hat{x}_2=\left(N_{22}-N_{12}^{T}N_{11}^{-1}N_{12}\right)^{-1}\left(C_2-N_{12}^{T}N_{11}^{-1}C_1\right) \tag{3.19}$$

$$\hat{x}_1=(N_{11}^{-1}C_1-N_{11}^{-1}N_{12}^{T}\hat{x}_2)$$

Once the unknowns have been estimated according to the required precision, the corrected values are used in the next phase. In the following steps, the matrix P contains a higher weight for the observations already compensated, since they are considered like a constraint. Such tie points, which have been already observed, could even be treated in order to reconstruct an adaptive mesh in support of the following positioning predictions.

Since the matrices involved in the computation are composed by blocks, they have been considered separately to speed up the computational operations.

At present, the software requires some improvements to work in real-time; just a few laboratory tests have been performed to check the functioning of some operative steps, such as the target recognition. Due to the Covid-19 pandemic, at the moment it is not possible to carry out further tests in collaboration with the FGI.

3.3. PLANNING AND CHECK: MENTION TO U.PH.O.

As already stated, U.Ph.O. is a tool for planning and checking the accuracy of a photogrammetric survey campaign. It has been mostly implemented by Eng. Daniele Passoni, during his PhD thesis work, in MATLAB® environment. In this context, a quick resume of his work is reported in order to frame the basic principles and the relevance of the adding module; for further information about the core of the procedure, it is recommended to deepen his own PhD thesis (Passoni, 2019).

The understanding of the metric potentialities of a photogrammetric survey is strictly related on the one hand to the recognition of its capability in building a numeric reconstruction of the observed objects, while on the other hand to the determination of its precision, both a-priori and a-posteriori.

The validation of the survey depends on the check of the achieved precisions a-posteriori, in order to provide the quality and the applicability limits of this information. Thus, the knowledge of the reachable precisions a-priori is necessary to plan the survey and its conditions, ensuring the expected precisions of the survey will be met. Nevertheless, none mission planner is able to perform this estimation, which is possible by means of rigorous tools for experts like CALGE (Forlani, 1986).

The novelty of U.Ph.O. consists in conceiving an instrument available also for medium-skilled users, more intuitive thanks to the ready-made graphical outputs and easily customizable for the specific working conditions. It permits on the one hand to obtain a rigorous evaluation of the expected precisions for a photogrammetric survey by UAV during its planning phase (a-priori), and on the other hand to check the goodness of the survey itself immediately at the end of the survey (a-posteriori).

3.3.1. Network design

The theory of evaluation of precisions is a relevant topic in the scientific literature. Two main approaches are used: the model of classical Photogrammetry and the terrestrial close-range one. According to Fraser (1984), accuracy is influenced by both the quality of the observations,

corresponding to the measurement of the coordinates of the points on the images, and the geometry of the network, called "network design". Network design analysis in engineering surveying are treated in Alberda (1980) and Niemeier (1982), while reliability of observations in close-range photogrammetric networks has been addressed by Grün (1978, 1980) and Torlegård (1980).

A classification of network design issues has been identified by Grafarend (1974):

- Zero-Order Design (ZOD): the datum problem; it is solved by means of the GCPs coordinates.
- First Order Design (FOD): the configuration problem; it involves the choice of an appropriate shooting geometry for a given array of object target points.
- Second Order design (SOD): the weight problem; it involves only an optimization of the scalar value σ, since each observed coordinate has the same accuracy.
- Third Order Design (TOD): the densification problem; since object point precision is largely independent of the target array density in photogrammetric survey with "strong" geometries, the densification problem does not seem to arise for the photogrammetric UAV survey.

The datum problem (ZOD) is solved by the introduction of the coordinates of GCPs, whose number is equal or greater than the rank deficiency to fix the RS. The problems of densification (TOD) and weight (SOD) are less influencing than the geometrical configuration (FOD). Therefore, the configuration problem FOD might be deepened in order to provide practical instruments for the planning.

While the characteristics of aerial blocks are well-known with nadir approach and standard overlapping, UAV blocks are usually characterized by the coexistence of oblique and nadir imagery, high overlaps and multiple cameras and sensors that require a specific approach to design the survey.

3.3.2. Classical Photogrammetry approach

Krauss introduces the estimation of expected accuracies in the traditional case of stereo restitution. This classical case is based on the assumption that the camera axes are parallel and normal to the base joining the cameras. This condition is always approximately fulfilled in aerial photogrammetry flights, where cameras are mounted on stabilized platforms; moreover, in case deviations from the ideal condition are consistent, the flight must be repeated.

Thus, it is possible to compute the object coordinates (X, Y, Z) from the quantities measured on the image and to evaluate the accuracy of these indirectly derived coordinates. The standard deviation along the Z axis is generally the most significant one. Supposing the principal distance c and baseline B error-free, the precision (standard deviation) of the elevation σ_z could be obtained by (1.20).

In the following Figure 3-8, the procedure to evaluate the expected precision in the stereo normal case is shown:

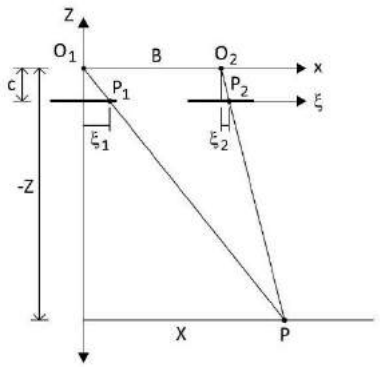


Figure 3-8 Stereo normal case, standard deviation evaluations.

Historically, the production of cartography adopted the normal case of stereo restitution as the most practical and efficient; hence, it ignores the possibility of convergent imagery as well as of the visibility of an object point in more than two images. This was valid for the classical analytical photogrammetry, which typically used to fly with a (60-70)% forward and (20-40)% transverse overlapping. As already mentioned, the use of digital cameras, UAVs and SFM software packages lead to change the methods of photogrammetric shooting and the conditions of stereo normal cases are rarely met.

3.3.3. The Terrestrial Close-Range Model

The relations used for the terrestrial close-range photogrammetry are significantly different because of the high level of overlapping and the presence of oblique images. In facts, the number of images and their convergence are fundamental aspects to take into account for the computation of the final accuracy. According to Fraser (1992), the precision of an object point (X, Y, Z) can be expressed as:

$$\sigma_{XYZ} = \frac{qZ}{c\sqrt{k}}\,\sigma_{p\xi} \tag{3.20}$$

Z is the distance between the object and the camera, c is the focal length, k is the number of overlapping images, q is a form factor and $\sigma_{p\xi}$ is the precision of the collimations on the images. The last parameter $\sigma_{p\xi}$ is strongly dependent by the measurement principle used and by the quality of the image itself. In case of calibrated non-metric cameras, it can be assumed equal to the pixel size. This model of accuracy conforms better to the conditions of a photogrammetric survey by UAV, since it takes into account the strong overlap of images and the shooting geometry, which is rarely nadir and often proves to be convergent, by means of the introduction of the parameters k and q.

However, the form factor q could be estimated only by empirical analysis (Fraser, 1984; Pagliari et al., 2017). From literature, it could be considered equal to 3.5 in case of nadir acquisition with a standard overlapping equal to 60% of the image size, 3 in case of acquisition with high cross overlapping and 0.4 in case of highly convergent geometry. The choice of the q

factor is dependent by user experience and directly influence the evaluation of the expected precision.

3.3.4. The Rigorous Approach principles

The application of a-priori expected precisions estimation methods, developed and used in aerial and close-range photogrammetry, gives approximate results not always reliable for the UAV case. In facts, on the one hand the estimation of expected accuracy used by Krauss excessively simplifies the imaging geometry and does not take into account the redundancy of the observation of tie points. On the other hand, the evaluation method introduced by Fraser for close-range photogrammetry requires the knowledge of the factor q, difficult to estimate for complex shooting geometries, like in a survey by UAV. Moreover, these methods do not take into account the uncertainties in the propagation of variance due to the process of orientation of the photogrammetric block itself.

A least squares approach, based on network simulation, is proposed as alternative. The positions and attitudes of the camera together with its calibration info and a Digital Surface Model (DSM) of the ground/object represent the inputs for the procedure, consisting in an optical triangulation technique. It is possible to determine one by one what portion is seen or not from the various points of view and the achievable precision, starting from the shape of the analysed area. The tool own a module for planning and another for checking after the survey campaign.

Bundle Adjustment, i.e., a non-linear least-squares method, is applied in order to estimate the unknown parameters together with their corresponding standard deviation.

Applying the least-squares approach, the resulting linear functional and stochastic model is:

$$v=Ax-l \tag{3.21}$$

$$C_{ll}=\sigma_0^2 P^{-1} \tag{3.22}$$

where l, v, and x are the vectors of observations, residuals and unknown parameters, respectively; A is the design matrix; C_{ll} is the covariance matrix of observations; P is the weight matrix; σ_0 is the variance factor.

The vector of estimated parameters is calculated as:

$$\hat{x}=(A^T PA)^{-1} A^T P\, l=Q_x A^T Pl \tag{3.23}$$

while the covariance matrix of the unknown parameters C_{xx}, which provides the variances, thus, the accuracy of the object coordinates, results being:

$$C_{xx}=\sigma_0^2 (A^T PA)^{-1} \tag{3.24}$$

In the analysed scenario, the fundamental observable is the image coordinate pair (ξ, η) on the frame.

In addition, GCPs coordinates are provided in order to properly lock the necessary degrees of freedom, thus, removing the rank deficiency due to the lack of constraints.

The resulting design matrix A is made by blocks, as described in the Table 3-5.

Table 3-5 Blocks of design matrix A.

	3×nr observed points	6×nr frames
2×nr observations	coefficients of the unknowns of the observed points	coefficients of the EO parameters
6×nr frames (pseudo-observations)		constraints coefficients of the EO parameters
3×nr GCPs (pseudo-observations)	constraints coefficients of the GCPs coordinates	

Let's suppose to know the stochastic model of observations, represented by the matrix $C_{yy} = \sigma_0^2 P^{-1}$. The weights of the constraints reported in this matrix could vary according to the conditions of the survey. Referring in particular to the EO, typically the constraints on the coordinates of the projection centres collected by means of an on-board GPS on the UAV allow an accuracy of about 10 m, due to the code solution. Conversely, using a double-frequency or RTK on-board GPS the obtainable accuracy reaches the order of 3 cm.

$$C_{yy}^{-1}=\begin{vmatrix} \frac{1}{\sigma^2_{obs}} & & & & \\ & \ddots & & & \\ & & \frac{1}{\sigma^2_{XYZ\ proj\ centres}} & & \\ & & & \ddots & \\ & & & & \frac{1}{\sigma^2_{XYZ\ GCPs}} \end{vmatrix} \quad (3.25)$$

The so-obtained C_{yy}^{-1}, together with the design matrix A previously presented, allows the calculation of the normal matrix N by means of the following relation:

$$N=A^T \cdot C_{yy}^{-1} \cdot A \quad (3.26)$$

Inverting the normal matrix N and extracting the main diagonal of the inverse for such indexes pertaining the X, Y, Z coordinates of the voxels, a vector containing the variance σ^2 associated to these unknowns is generated. Hence, applying the square root and reorganising the values distinguishing them according to the reference X, Y, Z component and to the corresponding voxel, the final accuracy maps based on the standard deviation σ are built.

This is the starting point to obtain a least squares simulation, giving as input to the procedure the IO and EO parameters, together with an approximate DSM of the area. The accuracy for each point visible from at least two frames could be estimated, starting from its effective visibility obtained by taking into account the obstructions automatically inferred from the DSM.

3.3.5. Adding module to deal with vertical façades

The main contribution in this work has been represented by a function intended to properly treat vertical walls. Starting from DSM, the global minimum and maximum are identified, thus, 3D voxels are created below the height value related to the single cell. If the obtained voxels are hidden by the surrounding context, they are not taken into account in the processing; otherwise, the voxels that result visible from images despite they stay below the surface level are useful to recognize the façades. The management of vertical walls is particularly useful in presence of convergent images with attitudes significantly different from the nadir one.

The procedure approach is maintained equal to the one established by Eng. Daniele Passoni, using as first the DSM as input to check visibility. Thereafter, in order to check the visibility in presence of hypothetic vertical walls, the analysis is repeated considering the layers below the top surface, according to the voxel subdivision in height of the same size of planimetric cells, from the top (DSM cell value) to the lowest level (global minimum of DSM). Thus, the original DSM

is modified step by step only in the specific cell, replacing the original height with the one of the analysed layer, in order to be given as input for the function *los2* in MATLAB© inferring correctly the potential occlusions from the surrounding environment. Moreover, in this regard, whenever an intermediate layer is reached by a collinearity ray from above, the correspondence is not considered as complied, since the stacking upper cell temporarily excluded represent an obstruction to the real visibility. In order to check this condition, once the visibility of the cell is confirmed by using the *los2* function on the modified DSM, the position of the intersection between the collinearity ray and the upper plane of the voxel, containing the top face of the cube, is verified. If it falls out of the top square limits of the cube, it is excluded that the collinearity ray pass through the "shaded" voxel standing above.

Hence, the updated least-squares approach considers as unknowns all the coordinates of such voxels that pass the test of visibility, together with the EO parameters of the images that see them. The observations are the ξ and η image coordinates of each pixel matching the seen voxels.

The design matrix is properly built in order to easily manage the indexes aiming to restore the voxel layer subdivision in the end of the processing, returning as output the accuracy maps per each Z interval.

TRAINING TESTS AND VALIDATION

The implemented tools have been tested on several case studies characterized by different complex scenarios, including areas affected by earthquakes and unsafe structures.

4.1. MAGO

MAGO has been tested in several circumstances during its development. The preliminary version, hereafter v.0, is preceding the modifications described in paragraph 3.1.3 and following. It has been evaluated reconstructing the orthophotos of a box and a structural lesion on a bridge arch. The version v.1 has been tested on a dataset collected in the Pompeii Archaeological Park, representing the scenario of *Regio* VII, *Insula* 14, in order to specifically evaluate the potentialities of the module for the automation in "unrolled" joint orthophotos reconstruction. In addition, the previous case studies have been repeated in order to compare the results highlighting the improvements related to the adding modules previously described at paragraphs 3.1.4, 3.1.5 and 3.1.6. Finally, a mention is due to the application of the semantic segmentation function introduced in paragraph 3.1.7.

4.1.1. Laboratory test on a box

A simple laboratory case study has been provided by means of a box over a checkerboard (Gagliolo et al., 2019b). Three nadir images have been acquired with a Canon EOS 40D camera with a focal length of 22 mm, at a distance of about 1.10 m. The resulting GSD is 0.3 mm. The images have been processed using Agisoft Metashape© and the resulting dense cloud has been filtered from noise using CloudCompare. The final point cloud consists in about 4500 points, with a mean point spacing of 10 mm. The so-obtained point cloud has been directly processed with MAGO v.0, while it has been imported in Agisoft Metashape© and substituted to the dense cloud for the meshing and the orthophoto reconstruction. This guarantees the coherence of the input data in the two methods of orthophoto generation.

The grid cell dimension *s*, which in MAGO v.0 was superimposed by the user and not automatically computed by the software, has been fixed to a precautionary value of 14 mm, to be almost sure to find at least one point in the cell and to minimize the empty cells.

IO and EO have been exported from Agisoft Metashape© processing. The central image has been chosen for the orthophoto reconstruction, which is realized over the horizontal (XY) plane.

The resolution of the orthophoto has been set to 2 mm, so the theoretical downscaling factor results lower than 7, taking into account the ratio between the orthophoto resolution and the original GSD, as in the following equation:

$$\text{downscaling factor} = \frac{\text{orthophoto resolution}}{\text{original GSD}} = \frac{2}{0.3} < 7 \tag{4.1}$$

As a precaution, the final downscaling factor has been fixed to 3 in order to avoid an excessive loss of resolution in the input image and, consequently, in the orthophoto. Thus, the image downscaled GSD results 0.3 × 3 = 0.9 mm, which is significantly lower than the orthophoto pixel size.

The boundaries have been automatically acquired from the limits of the point cloud, which represents a small portion of the scene, as depicted in Figure 4-1.

Finally, the local RS has its origin in the bottom-left vertex of the left checkerboard, and it is oriented as follows:

- X axis along the bottom side, oriented to the right;
- Y axis along the left side, oriented to the top;
- Z axis orthogonal to the floor, oriented upward.

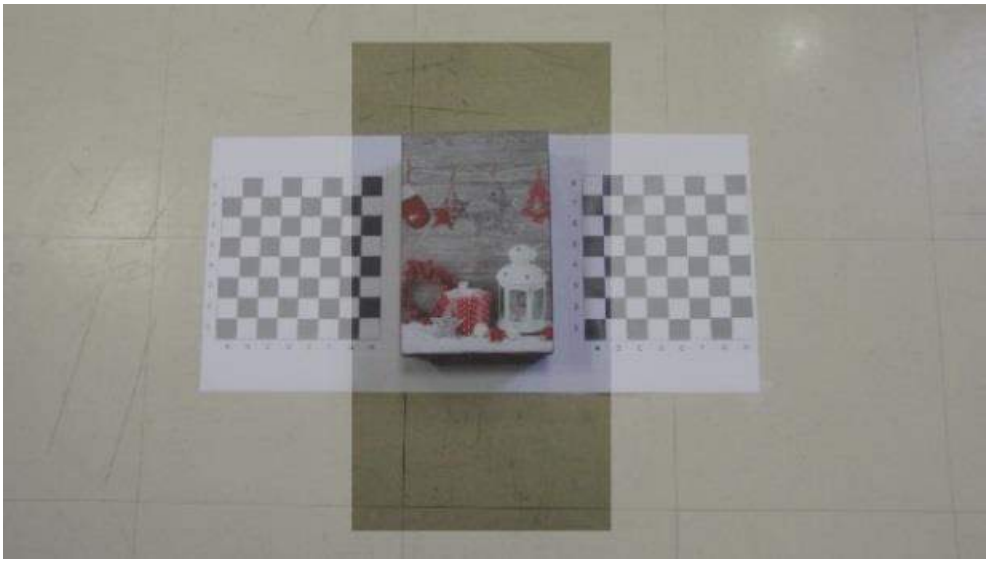


Figure 4-1 Analysed portion for orthophoto reconstruction.

As already mentioned, the most common existing photogrammetric software provide for orthophoto generation at a certain step of their workflow, typically after the generation of a mesh or a DSM from the point cloud. On the contrary, MAGO employs the point cloud to generate the orthophoto at its highest possible resolution, exploiting the original resolution of the images, which are projected over the adaptive mesh according to the IO and EO parameters. The resulting orthophoto is represented in Figure 4-2a, where the black areas correspond to no-match pixels. In Figure 4-2b, the DSM of the input point cloud highlights

that the holes in Figure 4-2a correspond to empty cells, due to border effects and to lack of data around the box.

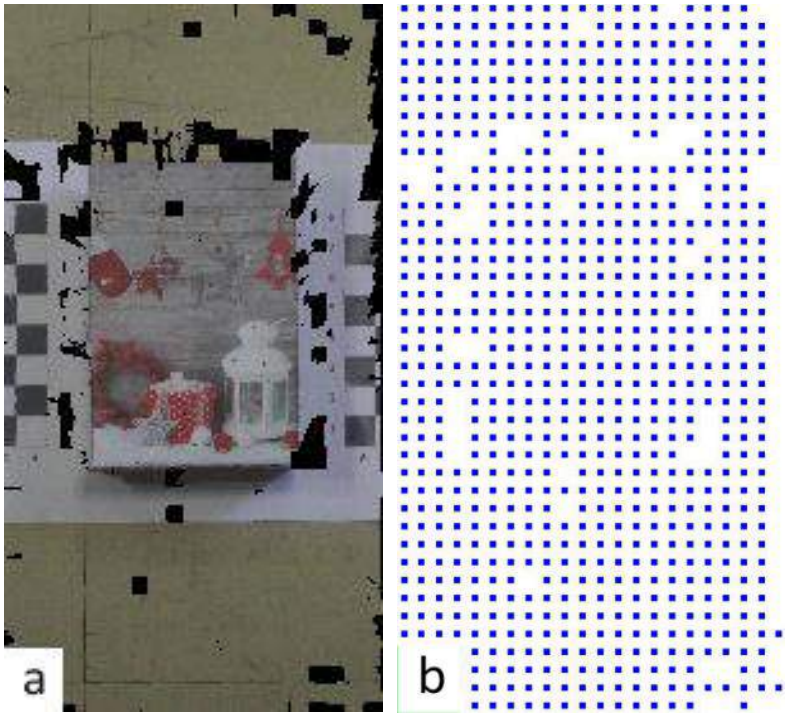


Figure 4-2. Orthophoto (a) and DSM (b) generated by MAGO v.0.

The orthophoto derived from MAGO v.0 and Agisoft Metashape© using the same settings, i.e., orthophoto plane, pixel size, orthophoto boundaries, processed image and input point cloud, are compared in the following.

The processing of the latter software has been set using the "Height field" surface type (which is the most suitable for planar objects, as the present case is) and the medium "Face count" for the mesh reconstruction, using the input filtered point cloud as input. The interpolation has been disabled, in order to reconstruct the mesh only where points are present. Then, the textured model has been realized, specifying to use only the central image and disabling the hole filling.

Considering the available options to build the orthophoto in Agisoft Metashape©, the pixel size is customizable. A default value, assumed as the highest theoretical value, is suggested according to the average GSD of the original images. Then, the dimensions of the orthophoto are automatically computed, based on the pixel and the mesh dimensions.

The orthophoto projection plane is chosen by means of three points (*markers*, representative of Ground Control Points, GCP), or defining a plane parallel to pre-defined views or to a user-defined current view.

The user may also choose the limits of the orthophoto by operating on the boundary settings, in case only a specific portion of the mesh should be represented in the orthophoto.

The orthophotos derived from MAGO v.0 and Agisoft Metashape© are depicted in Figure 4-3, where the black areas represent holes in the orthophoto reconstruction.

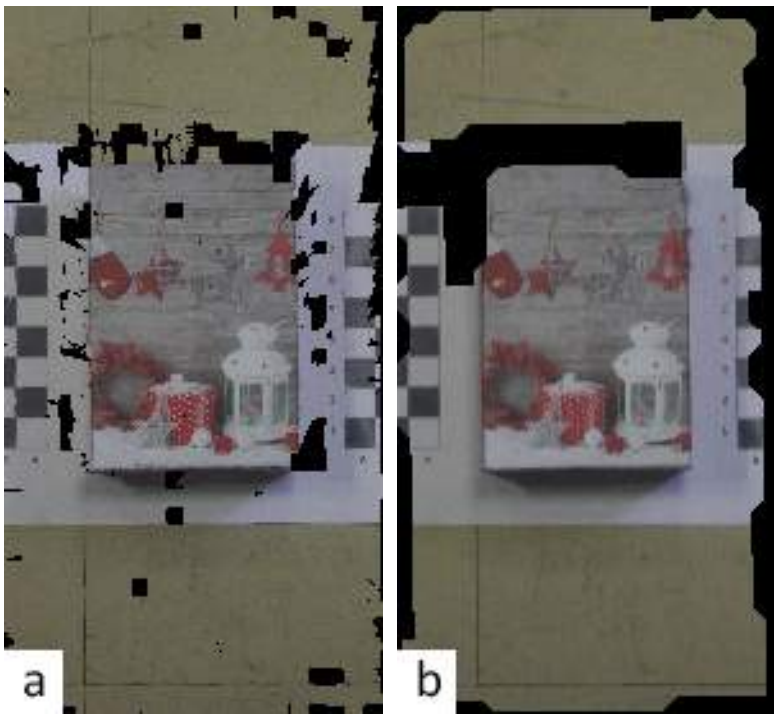


Figure 4-3. Orthophoto generated by MAGO v.0 (a) and by Agisoft Metashape© (b).

Figure 4-3 underlines similarities and differences in the outputs: both of them present holes around the box; MAGO v.0 produces more scattered and sparser no-data areas, whereas Agisoft Metashape© concentrates them around the box and along the image boundaries. Moreover, MAGO v.0 orthophoto seems clearer, with sharper transitions in colours between adjacent pixels.

In Figure 4-4, the orthophotos have been analysed in order to classify the holes generated only by MAGO v.0 (yellow), only by Agisoft Metashape© (cyan), and by both software (magenta). The background orthophoto is the one generated by MAGO v.0.

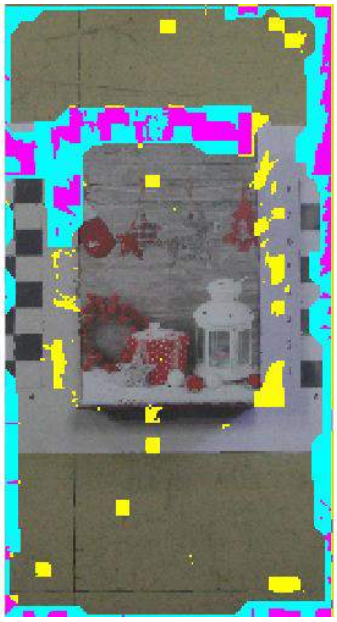


Figure 4-4. Comparison between holes in the orthophotos produced by MAGO v.0 and Agisoft Metashape©: yellow, cyan and magenta represent the holes produced by only MAGO v.0, only Agisoft Metashape©, and both, respectively.

Moreover, in Table 4-1 the number of cells and the percentage cover with respect to the orthophoto size (148 × 277 = 40996 cells) are reported. The number of no-data cells is obtained as follows: the yellow and cyan are the total hole cells produced by MAGO v.0 and Agisoft Metashape© respectively, having already removed the common ones (magenta).

The computation of the intersection area, the related statistics and the following elaborations have been performed using GRASS GIS 7.4 (GRASS Development Team, 2018).

Table 4-1 Comparison between MAGO v.0 and Agisoft Metashape© orthophoto holes: cell count and percentage cover data.

Orthophoto holes	**Cell count**	**Percentage cover**
MAGO v.0 only (yellow)	1563	3.81%
Agisoft Metashape© only (cyan)	4720	11.51%
Both (magenta)	1817	4.43%

Another statistical analysis has been performed on the three bands (Red, Green and Blue; hereafter R, G, B) of the two orthophotos. Considering each band separately, the previously individuated holes areas have been removed, assigning a no-data value. Then, the difference between MAGO v.0 and Agisoft Metashape© has been computed via a raster algebra

calculator, obtaining values in the range of about -100 to +100. The three differences maps have been classified according to the following ranges:

- class 1: difference values in the interval [-100;-20);
- class 2: difference values in the interval [-20;20];
- class 3: difference values in the interval (20;100].

The result of the classification is depicted in Figure 4-5 (a, b and c refer to R, G, and B bands respectively), where the grey areas correspond to class 2, whereas blue and red pixels lie into class 1 and 3, respectively.

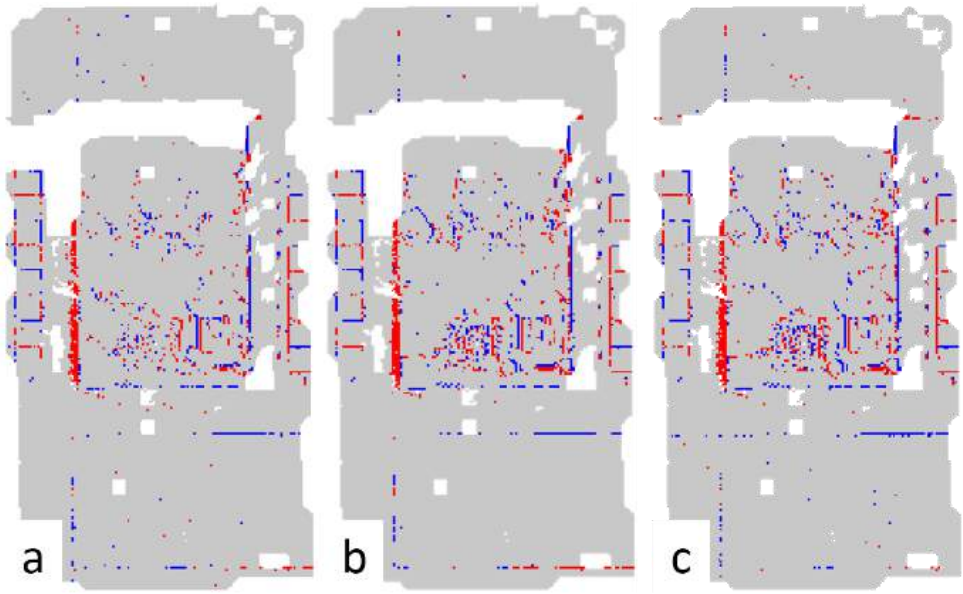


Figure 4-5. Classification of difference maps (MAGO v.0-Agisoft Metashape©) for R (a), G (b), and B (c) bands. Class 1, 2 and 3 are represented in blue, grey and red, respectively

Observing Figure 4-5, it is evident that the difference between the two orthophoto is limited between -20 and +20 for the majority of pixels (the ones represented in grey colour). The most marked differences are located along the borders of the box and the checkerboard and, in general, where a sharp change of colour in adjacent pixels is present, e.g., along the outlines of the floor tiles and of the squares on the checkerboard. Moreover, a more scattered pattern of high differences can be noticed over the box cover; again, it can be imputable to the change of colour between neighbouring pixels.

Table 4-2 summarizes the number of cells and the percentage cover of each class for the difference map, with analogous values for the three bands.

It should be noted that the total number of cells is 32743 for each band (instead of 148 × 277 = 40996), because the pixels corresponding to holes have been previously removed. The percentage covers are computed accordingly; indeed, the sum of percentage cover for each band is 100%.

Table 4-2 Classification of difference maps on the RGB bands: cell count and percentage cover data for the three classes.

	Class	Cell count	Percentage cover
Red band	1	536	1.64%
	2	31469	96.11%
	3	738	2.25%
Green band	1	633	1.93%
	2	31238	95.41%
	3	872	2.66%
Blue band	1	674	2.06%
	2	31209	95.32%
	3	860	2.62%

A final summary map of differences has been computed, as reported in Figure 4-6.

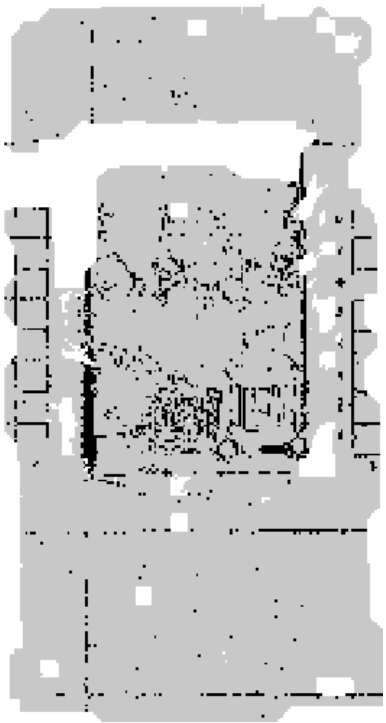


Figure 4-6. Summary of difference maps for the three bands. The grey and black pixels respectively represent areas of agreement and disagreement between the bands of the two orthophotos.

The grey pixels represent the areas where the R, G and B bands present a difference value between MAGO v.0 and Agisoft Metashape© within class 2, i.e., limited in the range of -20 and +20, whereas the black pixels correspond to areas where the considered pixel lies in class 1 or 3 for at least one of the three bands. Summarizing, the grey pixels can be interpreted as areas of limited differences for the three bands, namely areas where the two orthophotos are quite similar. Conversely, the black pixels highlight a disagreement of the different bands of the two orthophotos, and they could be due to the fact that MAGO v.0 does not any average among input pixels to attribute the value to the orthophoto pixel.

The same case study has been further analysed under the same conditions using the updated MAGO v.1. The final result is depicted in Figure 4-7, while the statistics related to the comparison between MAGO v.0 and v.1 are reported in Table 4-3 where the orthophoto holes are counted per each version.

Table 4-3 Comparison between MAGO v.0 and v.1 orthophoto holes: cell count and percentage cover data.

Orthophoto holes	**Cell count**	**Percentage cover**
MAGO v.0	3380	8.24%
MAGO v.1	2591	6.32%

Figure 4-7. MAGO v.1 orthophoto.

Table 4-3 shows the significant reduction of the number of orthophoto holes in MAGO v.1 in respect of v.0 (of the order of 23%). Comparing Figures 4-4 and 4-7, it is worth noting that the main improvements regard the isolated holes filling, the removal of mismatches around the box borders, together with the higher computational efficiency. In facts, the orthophoto production took just six seconds, despite the duration of a few minutes in the v.0.

4.1.2. Supporting the structural evaluation of a lesion in a bridge arch

MAGO workflow has been even employed in supporting the structural evaluation of a lesion in a bridge arch (Figure 4-8), located in the Genoese countryside (Gagliolo, 2019c). Despite it is a small structure, its strategic relevance is due to the fact that it is the only connecting route between the paths alongshore and hilly.

Since several years, the development of the structural lesions was raising concerns, thus, periodical monitoring activities were fulfilled and, recently, retrofitting operations have been achieved.

Figure 4-8 Valley arch of the Bridge.

The point cloud taken into account for the application of MAGO procedure have been obtained from an integrated survey of the structure, acquired by means of TLS, Photogrammetry and Total Station (TS), which has been used to achieve the topographic survey of the GCPs.

The employed instrumentation includes the camera Canon EOS 40D, the TLS Z+F Imager® 5006h and the TS Leica TCR703.

The portion highlighted in red in Figure 4-9 has been processed in MAGO v.0; it represents the upper part of the main lesion, located in the valley arch of the bridge.

Figure 4-9 Image projected for the orthophoto reconstruction; the portion of interest is highlighted in red.

Two orthophotos have been generated, starting from both the TLS and the photogrammetric point clouds. The two products have been compared in order to study their differences.

The GSD of the considered image is about 2.5 mm (obtained by a focal distance of 85 mm and a distance from the object of 35 m), while the mean spacing between points of the clouds is 4 mm for TLS and 6 mm for Photogrammetry, respectively. The resolution of the orthophoto has been set in both the cases at 4 mm, comparable with the TLS point cloud resolution and enough to appreciate the lesion of interest. The original image has been processed at the original resolution, without the application of a downscaling factor.

The parameter *s*, corresponding to the grid spacing, which is not automatically computed in MAGO v.0, has been fixed to 16 mm for the photogrammetric case study and 12 mm for the TLS one, respectively. This oversizing is precautionary since it allows to detect the real lack of data where there are no points in the grid discretization.

The needed photogrammetric processing has been achieved with the commercial software Agisoft Photoscan©, while the point clouds management has been carried out by means of CloudCompare.

The resulting model has been properly rotated in order to put the analysed front parallel to the XY plane, since in MAGO v.0 it was not yet possible to freely choose the orthophoto plane.

Orthophotos allows both the metrical and thematic investigation of the object, while for the point clouds is not always possible to obtain both features in a satisfying way.

In particular, the TLS point cloud (Figure 4-10) guarantee a better geometric resolution than the photogrammetric one, but the grey scale employed for the radiometric representation does not allow the easy recognition of objects. On the other hand, the photogrammetric point cloud (Figure 4-11) permits the easy recognition of objects thanks to the colour, but the obtainable resolution is typically more rough.

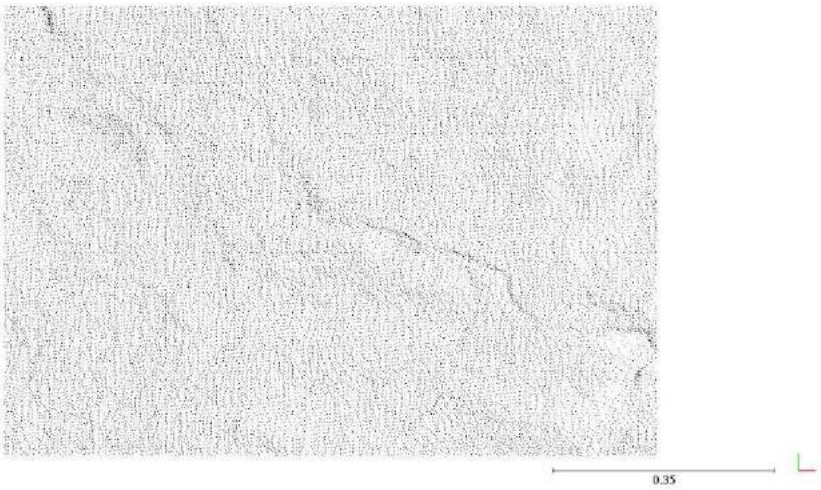


Figure 4-10 TLS point cloud.

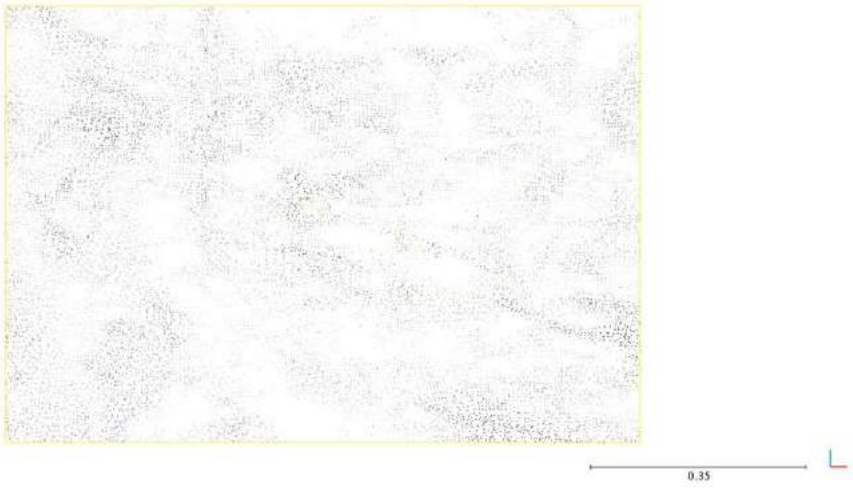


Figure 4-11 Photogrammetric point cloud.

The orthophoto obtained by MAGO allows to generate from both the input point clouds a high resolution product.

The resulting orthophotos from TLS and Photogrammetry respectively are reported in Figure 4-12 and Figure 4-13. The portions represented in deep black in both these figures are due to the lack of data, thus the spacing between point is higher than the grid resolution, or to the missing direct correspondence for the projection. These bugs are partially solved in MAGO v.1.

Figure 4-12 Orthophoto obtained by TLS point cloud.

Figure 4-13 Orthophoto obtained by photogrammetric point cloud.

An unexpected instance is the presence of a visibly higher lacks of data in the orthophoto coming from the TLS point cloud, although the higher mean density than the photogrammetric one.

In order to analyse this behaviour, the two DSMs reported in Figure 4-14 and Figure 4-15 have been examined; empty cells are identified by red dots, while in light blue are represented the filled ones.

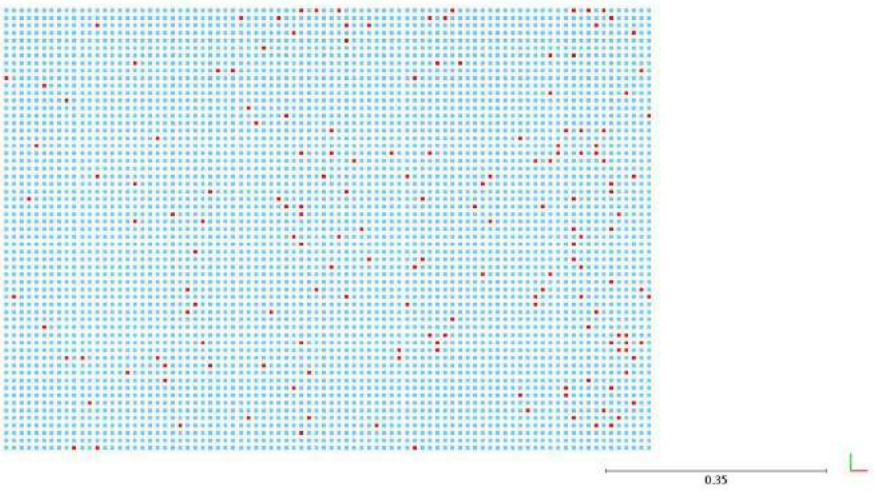


Figure 4-14 TLS DSM; cell size 12 mm. The red points represent the isolated holes.

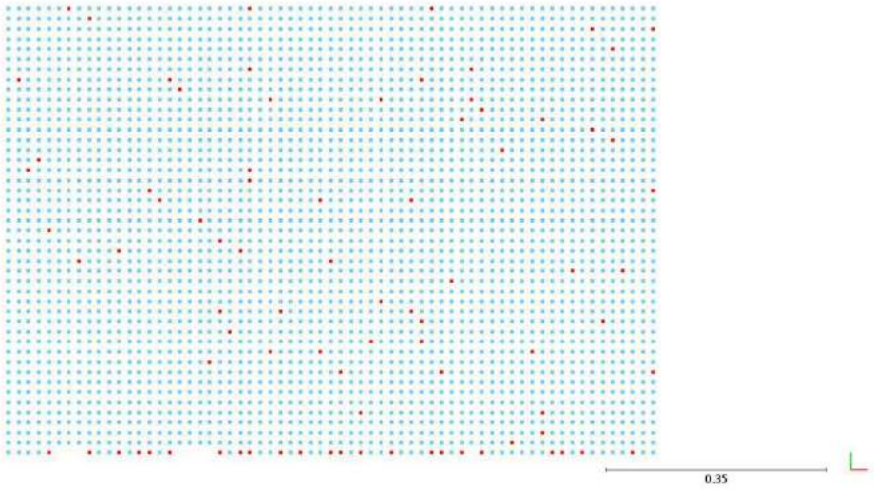


Figure 4-15 Photogrammetric DSM; cell size 16 mm. The red points represent the isolated holes.

Despite the choice of the grid size *s* has been made precautionary for both the case studies, the TLS DSM highlights a higher quantity of isolated holes. This fact indicates the irregularity of the points distribution, not enough compensated by the creation of a grid with a size highly larger than the mean spacing between points.

In Figure 4-16, described in Table 4-4, the statistics about the portions with lack of data in the orthophotos are reported. By means of the GIS Open Source software GRASS, the size of the non-reconstructed areas has been computed, as well as their overlapping between the two orthophotos.

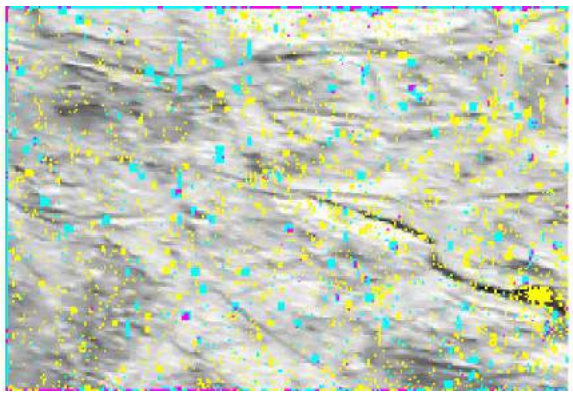


Figure 4-16 Identification of the non-reconstructed areas; in yellow the ones from TLS, in cyan the ones by Photogrammetry, in magenta the overlapping between the two techniques.

Table 4-4 Overlapping percentage of the non-reconstructed areas and number of cells in correspondence of the whole orthophoto (tot) or of the only lesion (les).

Non-reconstructed areas	Nr. cells		%	
	tot	les	tot	les
TLS only (yellow)	3431	173	7.5	31.6
Photogrammetry only (cyan)	1606	6	3.5	1.1
Both (magenta)	340	3	0.7	0.5

The statistics reported prove what has been visually evaluated. In facts, the areas with lack of data in the TLS orthophoto are about the double of the ones in the photogrammetric reconstruction. The common areas are present in a low percentage and mostly located in the borders proximity.

Although the input data have been provided to obtain comparable conditions, the significant dissimilarity lead to consider the different nature of the treated point clouds, rather than the applied procedure.

By repeating the same procedure using MAGO v.1, it is possible to remark that the estimation of the grid cell size performed using the v.0 is fully coherent with the automatic computation realized with the specific module, which returns 16 mm for the photogrammetric point cloud and 14 mm for the TLS one.

Moreover, the enhancement of the computational quality in MAGO v.1 allowed the processing of the whole façade, instead of the limitation to the single area of interest.

In Figure 4-17 and Figure 4-18, the orthophoto obtained by using MAGO v.1 for the case study with photogrammetric source cloud, resulted as the best in the previous comparison, is reported, representing both the whole arch and the single portion of interest analysed with MAGO v.0. In Table 4-5, analogously to Table 4-4, the statistics are shown with respect to the new case study, comparing it with the orthophoto generated by MAGO v.0 for the photogrammetric sample.

Table 4-5 Overlapping percentage of the non-reconstructed areas and number of cells in correspondence of the whole sample (tot) or of the only lesion (les) for the photogrammetric test case.

Non-reconstructed areas	Nr. cells		%	
	tot	**les**	**tot**	**les**
Photogrammetry v.0	1946	9	4.2	1.6
Photogrammetry v.1	883	11	1.9	2.0

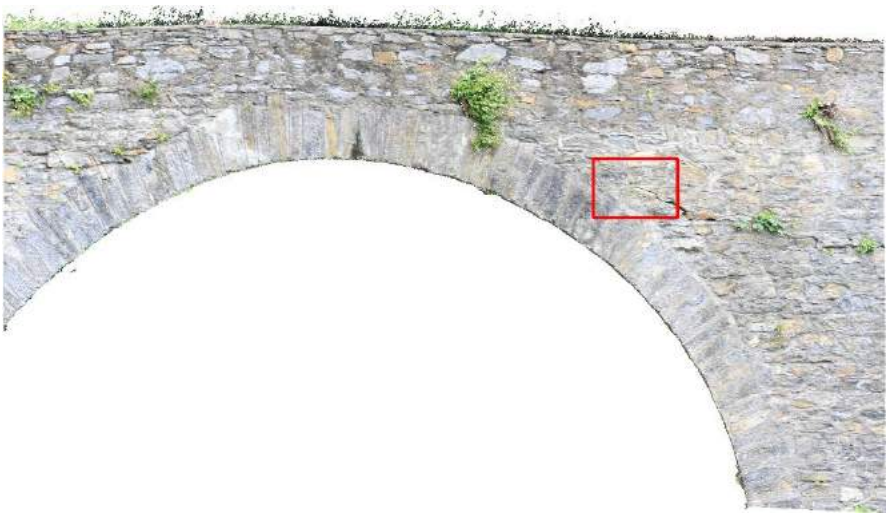


Figure 4-17 Orthophoto of the arch generated with MAGO v.1. In the red rectangle, the portion of interest of the previous tests is highlighted.

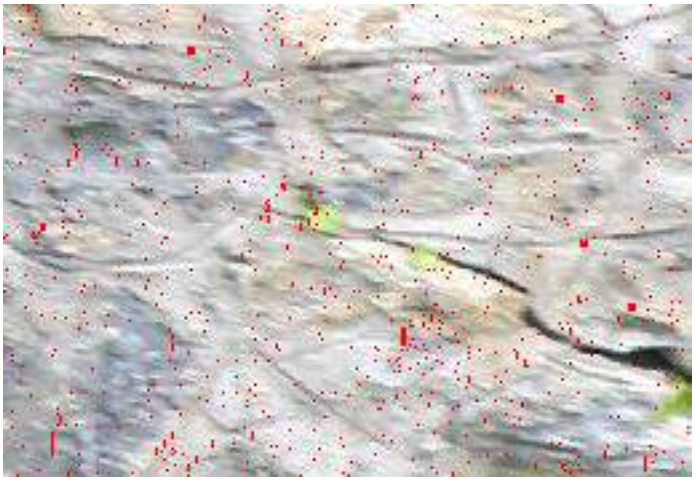


Figure 4-18 Focus on the specific sample generated by MAGO v.1; empty cells are highlighted in red.

An opposite behaviour is attributable to the whole sample or to the focus on the lesion. In facts, on the one hand the empty cells obtained with MAGO v.1 for the overall portion of

analysis are more than halved than MAGO v.0, while on the other hand the numerousness of the ones belonging to the single lesion is fully comparable, or rather a few higher. The latter condition is incoherent with the expectations; however, the majority of these cells are located on the border of the lesion, suggesting the affecting by boundary effects.

Finally, the clearest advantage is the possibility to treat the whole arch rather than the limited area focused on the specific lesion, allowing the analysis of the context in its entirety.

4.1.3. Historical and structural analysis of *Regio* VII, *Insula* 14 in the Archaeological Park of Pompeii

MAGO v.1, fully implemented with the inclusion of the module for the automatic adaptation of the plane and the "unrolling" of adjacent walls (paragraph 3.1.6), has been applied to the case study of *Regio* VII, *Insula* 14, a site located in the Archaeological Park of Pompeii (Capobianco et al., 2021). The interest towards this case study is born from the collaboration with an interdisciplinary research team of the Genoa University, including skills about archaeology, structural engineering and geomatics. The production of orthophotos is considered particularly useful in order to achieve further analysis about the actual state of conservation of the structures, examining the eventual presence of lesions on the walls and, then, establishing the potential necessity of retrofitting interventions. Moreover, the complete survey of the site will support the historical analysis about the relationship among the neighbour spaces.

Regio VII, *Insula* 14 (Figure 4-19) is a district of about (60×35) m, located in the western part of Pompeii: it faces south on the most important road artery of the city, *Via dell'Abbondanza*, and is bordered westward by *Vico della Maschera*, northward by *Vico dello Scheletro* and eastward by *Vico del Lupanare*. The area of interest is located close to some of the most important monuments of Pompeii, including the Civil Forum, the Stabian Baths, the Triangular Forum and the Theater District, confirming its importance in the urban development framework urban development. It includes three houses and 12 shops, divided among 20 entrances located on all four sides of the district.

The activities of the University of Genoa, started in 2016 and still ongoing, aim to reconstruct a detailed picture of the history of this district of Pompeii, by outlining its evolution over time, its phases of life, its relationship with the rest of the city, the function of the shops facing the street. In parallel to the excavation activity, which led to important results regarding the history and the function of the shops, a stratigraphic analysis of the walls was carried out, involving all the structures of the district. The aim of this activity is to highlight the construction events of the insula and its changes from a structural point of view, proposing a reconstruction of its appearance in some phases: traces of secondary elements (stairs, regular rows of holes in the walls, drain pipes) may suggest the presence of one or more upper floors. The survey and the structural analysis activities could confirm the hypotheses proposed by the stratigraphic analysis.

A survey campaign has been performed in the period 22nd-24th September 2020, using UAV and terrestrial Photogrammetry, TLS, TS and GNSS techniques. A preliminary inspection has been achieved onsite on 18th July 2020, in order to define the goals with the reunited research team, consisting in three groups expert in archaeology, geomatics and structural engineering.

The dataset employed to test MAGO v.1 consists of a sample of the photos collected by terrestrial photogrammetry during the preliminary inspection, which total corresponds to more than 900 images framing the external perimeter of the site. The considered chunk is made up of 80 photos, focused on a corner between two boundary façades (Figure 4-19). The post-processing of the data acquired both in this phase and during the proper survey campaign is still ongoing, due to the huge amount of collected images, which are a few less than 10000, including UAV and terrestrial photogrammetry. As already stated, the future aim will be the application of MAGO for the orthophoto production in order to support the static analysis of the entirety of the site.

Figure 4-19 Aerial view of the west part of the *Regio* VII, *Insula* 14 site. The area examined in MAGO v.1 processing is highlighted in the red rectangle in the bottom-left angle.

The camera used to collect the terrestrial images is a Canon EOS 40D, with a focal length of 17 mm. The resulting GSD at approximately 2 m of distance from the framed walls corresponds to 1 mm. The 80 images of the sample have been post-processed by means of the software Agisoft Metashape©, in order to obtain the input point cloud, characterized by a spacing between points of about 10 mm, together with the IO and EO parameters.

The orthophoto resolution has been set at 5 mm. The image chosen to be orthogonally projected on the orthophoto planes parallel to the two façades depicts clearly both sides, as shown in Figure 4-20.

Figure 4-20 Input image for the MAGO v.1 procedure.

In Figure 4-21, the resulting joint orthophoto of the two adjacent walls is depicted.

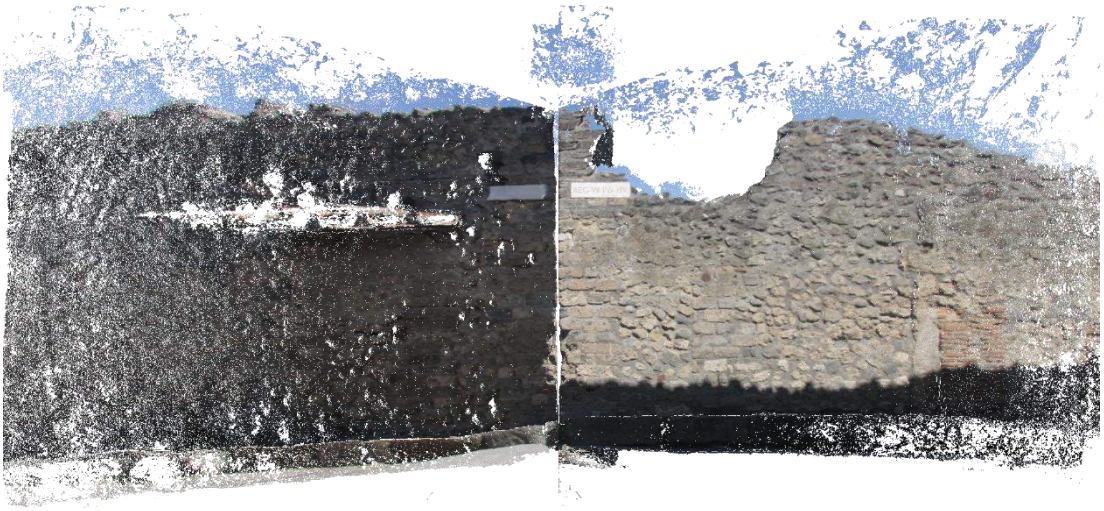


Figure 4-21 Resulting unrolled orthophoto of the adjacent façades.

It is worth noting that the source cloud is affected by noise due to the presence of unmasked sky all around the wall upper boundaries, which has been erroneously reconstructed by the photogrammetric software. The DSMs depicted in Figure 4-22 testify this condition.

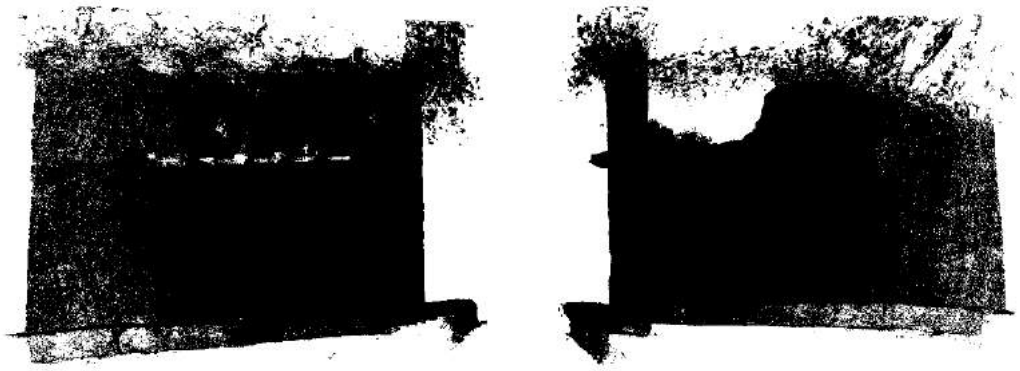


Figure 4-22 DSMs of the two adjacent sides.

A colour filter has been set within the procedure, in order to mask the sky in the orthophoto a-posteriori. A criterion similar to the one described in 3.1.7 for the semantic segmentation has been used, i.e., the image has been converted in HSV colour space and then masked, applying the *cv::inRange* function on the interval associated to the sky tone. The result is depicted in Figure 4-23.

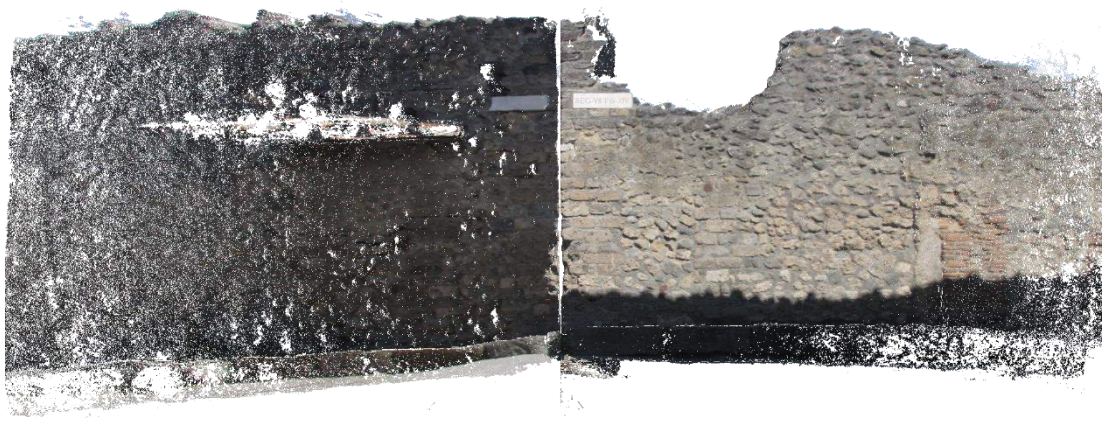


Figure 4-23 Resulting unrolled orthophoto of the adjacent façades with the application of the sky removal.

A further consideration is due with respect to the overall quality of the result. In facts, it is evident that the definition of the left façade is worse than the right one. In the present case, the spacing between points of the two walls is homogeneous, as inferable even from the DSM depicted in Figure 4-22, but the GSD of the employed image is significantly different due to the variation of the distance from the shortened shooting point. This problem could be avoided without increasing the orthophoto pixel size by using multiple images, consequently obtaining

an orthomosaic. As mentioned in 3.1.3, the implementation of such function as been sketched, thus, further tests will follow in order to validate its basic principles.

Of course, another influence on the overall quality is due to the different light exposure of the two sides, which significantly affects the result from both the radiometric and geometric point of view (Honkavaara et al., 2012).

4.1.4. Mention to the application of extra module for semantic segmentation on the UAVid case study

The analysed case study dedicated to the application of the extra module for semantic segmentation (paragraph 3.1.7) is represented by the UAVid dataset (Lyu et al., 2020), composed by high-resolution videos and imagery focusing on urban scenes, whose segmentation is based on eight object categories: buildings, roads, static cars, trees, low vegetation, humans, moving cars, and background clutter. The dataset consists of 42 video sequences (from "seq1" to "seq42"), which are captured with 4K high-resolution by the oblique point of view. The authors of the benchmark provided ten images extracted per each sequence, labelling the 420 resulting images with eight classes. Moreover, the sequences have been classified in three groups, i.e., training, test and validated sequences.

The proposed strategy exploit MAGO workflow for the production of geometric and radiometric maps, used as input for the procedure. Moreover, the final orthogonal projection of original images allows the comparison among consecutive epochs for the recognition of moving objects.

Since UAVid imagery is not exactly fitted for photogrammetric applications, the dataset employed as case study has been chosen paying attention that the image overlapping was sufficient to allow the 3D point cloud reconstruction by means of SFM technique.

Since the proposed approach requires the use of the 3D point cloud of the scene, the so-called "seq18" has been chosen as case study, given that the overlapping of provided frames was sufficient to obtain the photogrammetric reconstruction.

In Figure 4-24, the first epoch of the distributed "seq18" is shown.

Figure 4-24 First epoch of "seq18".

The 3D point cloud production has been achieved by means of Agisoft Metashape©. In this regard, the choice of the input information for the photogrammetric post-processing represented the first issue. In facts, on the one hand, the extracted images provided for each sequence are a few number compared to the usual photogrammetric blocks, and their resulting point cloud is affected by the presence of outliers due to the scarcity of the correspondences, as testified by Figure 4-25. On the other hand, the use of a higher number of frames (91) directly extracted from the provided videos has been attempted, leading to the production of a visibly distorted point cloud (Figure 4-26). This behaviour could be due to the fact that the drone path followed almost a straight line, as in a single strip dataset. Thus, the presence of a single strip, without any Ground Control Point to stabilize, has been badly managed by the software in the set conditions, causing a fleeting reconstruction.

Figure 4-25. Resulting point cloud from the provided frames of "seq18".

Figure 4-26. Lateral view of the point cloud obtained from 91 frames, with clearly visible distortions.

The following parameters have been set to carry out the workflow: the ten provided frames have been given as input, then the aerotriangulation has been performed at Medium quality, meaning that the image has been downscaled by factor of 4 (2 per each side). Finally, the dense cloud reconstruction has been launched at Ultra High quality, corresponding to process the images at their original resolution, and Mild depth filtering. Since no information about the Reference System was available, a pretended one has been attributed so that the building façades are vertical and the objects proportions are coherent with their standard measurements.

The resulting point cloud has been filtered using the Statistical Outlier Removal (SOR) and the noise filter available in CloudCompare, applying the suggested default parameters; moreover, it has been subsampled with a minimum spacing between points of 0.1 m.

The so-obtained 3D point cloud has been processed to obtain two DSMs, the former containing the geometric information in terms of Z coordinate median (Figure 4-27), while the latter containing the radiometric information converted to greyscale values (Figure 4-28). The following Figure 4-29 and Figure 4-30 are focused on a portion of interest, encircled in red in Figure 4-27 and Figure 4-28. This box, located where the obtained point cloud is sufficiently satisfying, has been chosen to carry out the test. The poor quality of the obtained point cloud for the analysed dataset is due to the fact that the shooting of the source images was not planned to obtain a 3D survey.

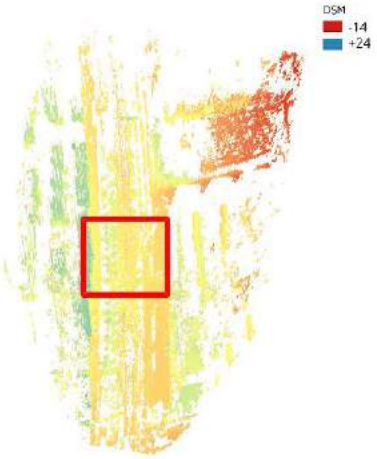


Figure 4-27. DSM with the highlighting of the focusing box.

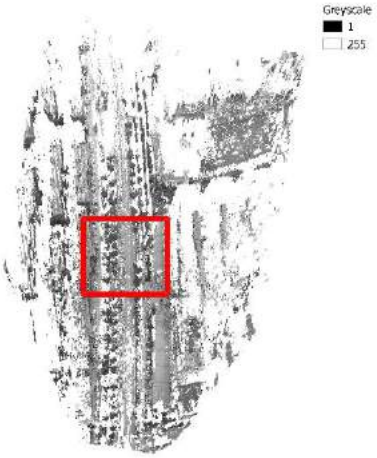


Figure 4-28. Corresponding greyscale map with the highlighting of the focusing box.

Figure 4-29. Detail of DSM focused on the indicated box.

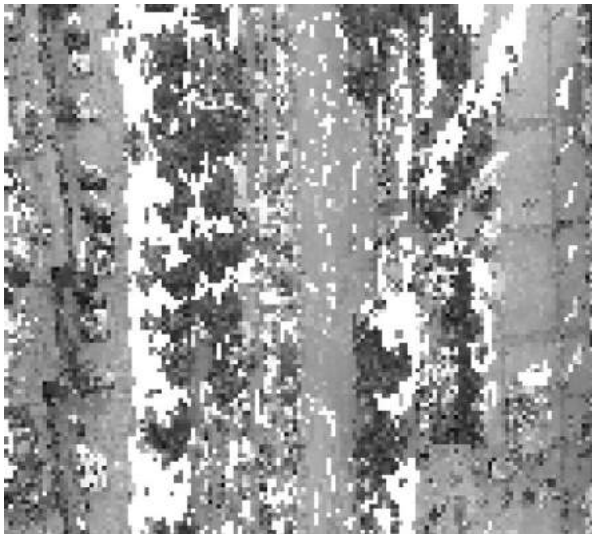


Figure 4-30. Detail of greyscale map focused on the indicated box.

Then, R_{class} map has been computed, as described in paragraph 3.1.7; Figure 4-31 shows the portion belonging to the established boundaries, indicated in Figure 4-27. It is possible noting that most of the areas are denoted with value 4, interpretable as high variability both in geometry and radiometry and both in the X and Y directions, or -2, identifying high variability in colour but null in geometry.

The starting R_{colour} and R_{geom} data are obtained using the *cv::cornerHarris* function, setting the kernel size at 7, comparable with the dimension of trees crowns and cars considering that each pixel is 0.5 m, the aperture parameter for the Sobel operator at 3, and the constant α at 0.04.

The R_{colour} and R_{geom} values considered as limit for the flat region (|R| small) are 10^{-5} and 10^{-9} respectively.

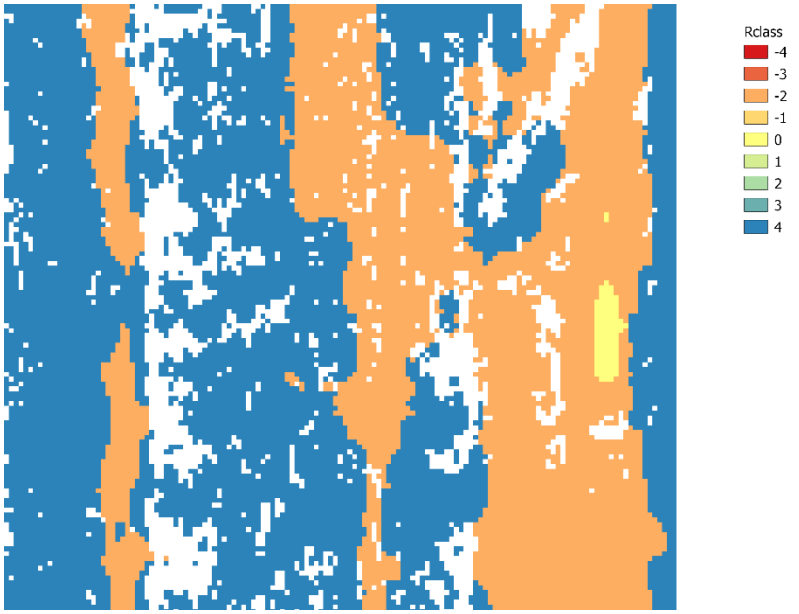


Figure 4-31 R_{class} map.

The preliminary segmentation, resulting from the sheer analysis of the static scenario, is depicted in Figure 4-32. It includes all the categories except for the moving objects (class 7).

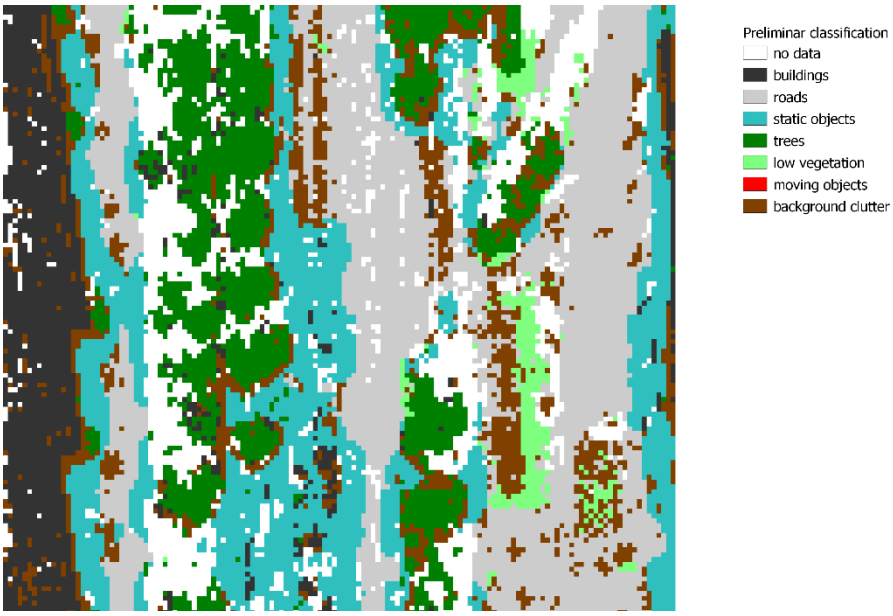


Figure 4-32 Preliminary segmentation of the static scenario.

The kernel used for the static objects research has size 5×5. It is worth noting that the class of the trees is satisfying, as well as the ones of buildings, roads and low vegetation. Regarding static objects, it is recognisable the presence of the cars waiting at the crosswalk or parked on the roadside.

Moving objects are isolated subtracting the resulting projections of consecutive epochs time by time, as shown in Figure 4-33. The points depicted in red represent the variations between two consecutive epochs. Nevertheless, their quantity is excessive with respect to the real presence of moving objects. The development of further strategy for the outlier removal would be needed in future. In facts, some mismatches could be associated to discontinuities, which are not perfectly overlapping even if pertaining to static elements.

From the shown results, the processing outcomes are strongly affected by the quality of the input data, i.e., the 3D point cloud and the deriving geometric and radiometric DSMs. Undoubtedly, the upstream 3D reconstruction is badly influenced by the presence of many moving objects, which would require a huge time to be singularly masked in each source frame in the Agisoft Metashape© interface. Thus, it could be worthy to bring forward the moving object detection, working already on their recognition on the 2D images, so that it would be possible to automatically exclude them from the point cloud reconstruction.

This proposal of further development, i.e., getting more information from the primary source represented by the acquired video and images, would allow treating moving objects as an opportunity to improve the whole workflow.

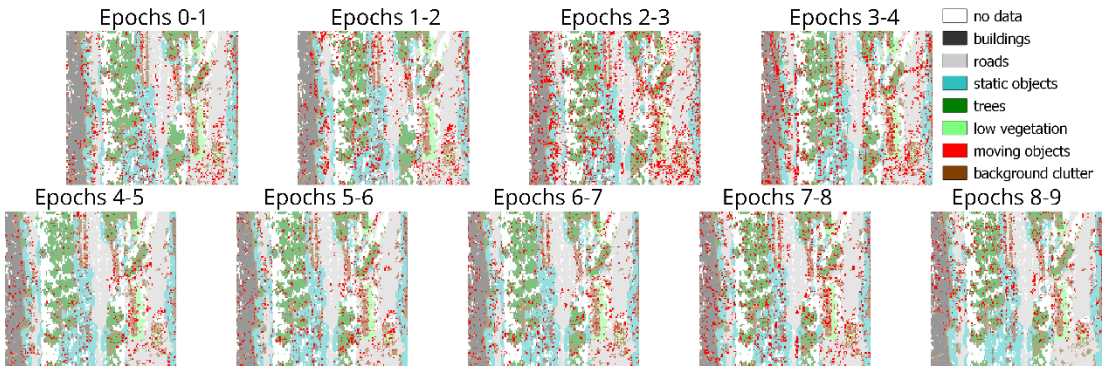


Figure 4-33 Final segmentation: moving objects varying in consecutive epochs are highlighted in red.

4.2. U.PH.O.

As already stated, the tool U.Ph.O. has been mainly developed by Eng. Daniele Passoni. Hence, two case studies are presented in order to show the achievements of his version and the further improvements obtained by means of the implementation of the module for the management of vertical walls. In particular, the couple of buildings made by Civic Tower (Norcia) and San Salvatore Church (Campi di Norcia) has been considered as test case for the

original version (hereafter v.a; Gagliolo et al., 2019a), while the update one (v.b) has been evaluated using the test case of *Regio* VII, *Insula* 14 in Pompeii.

4.2.1. Civic Tower (Norcia) and San Salvatore Church (Campi di Norcia)

The Civic Tower of Norcia and the San Salvatore Church of Campi di Norcia (Italy) were both strongly damaged by the earthquakes that took place in Central Italy in August and October 2016. Nevertheless, they were ascribable to two opposite scenarios, due to their location and state of conservation, as detailed in the following.

The Civic Tower stands out in San Benedetto Square, the core of the historical centre of Norcia. The structure is surrounded by other historical buildings, such as the Basilica of San Benedetto and the Co-Cathedral of Santa Maria Argentea, and private houses. Thus, the examined area is densely built-up, with possible occlusions and limitations for the flight.

Figure 4-34 depicts the Civic Tower and the façade of San Benedetto Basilica.

Figure 4-34 Civic Tower and the façade of San Benedetto Basilica.

Such a complex scenario for a UAV survey was further complicated by the severe safety restrictions applied in case of calamitous events, which strongly limit the direct access to the survey areas and so the flight operations. The aim of the securing operations carried out on the Civic Tower is the disassembly of the damaged structure, in order to avoid an incoherent collapse and then to retrofit and restore it brick-by-brick.

The survey campaign was performed in a single day (3rd November, 2016), followed by one day of photogrammetric post-processing (thanks to the optimized criteria reported in Gagliolo et al., 2017), which was necessary to design the reinforcing structures, realized on 5th November.

The survey campaign was carried out using a DJI© Inspire drone with an embedded Zenmuse X3 camera. The focal length is 3.61 mm, while the sensor size is 4000×3000 pixels with a pixel dimension of 1.56 μm. The flight was performed considering a nadir attitude and a relative height of about 50 m, which leads to a GSD of 22 mm. A support survey was performed using a GNSS receiver in order to acquire the GCPs coordinates.

The other analysed building, San Salvatore Church, is a small structure located in the countryside of Campi di Norcia. It was surrounded by a wide field, a parking area and it is adjacent to a cemetery.

The aim of the survey was a 3D description of the area, in particular of the church, where the inspection of the interior was important to document the presence of works of art and to design their recovery. The securing operations were particularly devoted to guarantee the safety during the removal of the works of art, and were not addressed to the conservation of the building, because of its highly compromised state, as shown in Figure 4-35.

Figure 4-35 San Salvatore Church ruins.

The survey campaign was performed in a single day (23rd November, 2016), using a DJI© Inspire drone with an embedded Zenmuse X5 camera. The focal length is 15 mm, while the sensor size is 4608×3456 pixels with a pixel dimension of 3.76 μm.

The flight was performed considering a nadir attitude and a relative height of about 50 m, which leads to a GSD of 13 mm.

As already mentioned for the Civic Tower, a GNSS survey was performed to acquire the GCPs coordinates.

The results of the application of U.Ph.O. on these scenarios, i.e., Civic Tower and San Salvatore Church, are presented and discussed in the following, reporting the occurrence map and the estimated precisions along three axes. The related statistics have been computed out using GRASS GIS, ver. 7.4.4 (2019), while the figures have been produced using QGIS, ver. 3.4 (QGIS Development Team, 2019). The coordinates are expressed in ETRF 2000-2008.0 reference frame, with UTM 32N projection.

In the former case, since the Civic Tower stands out against the surrounding buildings, it is not affected by the presence of obstructions, as inferable from the visibility map; conversely, the neighbouring areas result not visible from the planned shooting points because of the hiding due to the Tower itself.

The values in the occurrence map (Figure 4-36) correspond to the number of images that see each DSM cell, taking into account the presence of eventual occlusions. The black dots represent the resulting shooting points, obtained in the preliminary block design by U.Ph.O.

The average value of occurrence is 4.5, with a standard deviation of 1.8. The pattern of Figure 4-36 underlines what already conceivable from the geometry of both the Civic Tower and the shooting points: the cells with higher occurrence number are located in the centre of the survey area, whereas the occurrence number sharply decreases moving outwards.

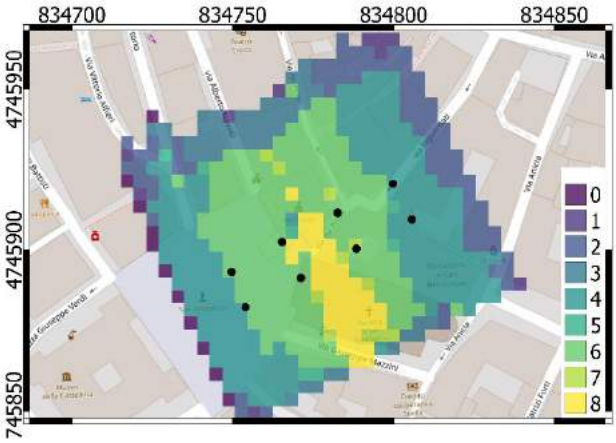


Figure 4-36 Occurrence map for the Civic Tower.

The occurrence map has been classified, according to the five classes reported in Table 4-6.

Table 4-6 Classification of occurrence map: cell count and percentage cover for the five classes.

Class	Occurrences nr.	Cell count	Percentage cover
0	0-1	21	3.6%
1	2	77	13.5%
2	3-4	235	41.1%
3	5-6	188	32.9%
4	7-8	51	8.9%

Class 1 corresponds to the minimum requirement of occurrences, i.e., two images, in order to obtain stereo-vision, needed condition to build a 3D model from images. For the pixels contained in class 0, it is impossible to reconstruct the point cloud representing the object. For classes 2-4, a 3D model could be successfully reconstructed, even with the possibility of control, due to the increasing redundancy related to higher class number.

Only a small amount (3.6%) of pixels lie in class 0; the majority of pixels falls in class 2, where 3 and 4 number of occurrences are contained, and a significant amount lie in class 3. Only 11 pixels are not seen (0 number of occurrences) and 40 are seen by all the eight shooting points.

From Figure 4-37 to Figure 4-39, the estimated precision along X, Y, Z axes are shown, respectively. The average values of standard deviations are 4.1, 4.6 and 8.5 cm, respectively. As expected, higher values characterize σ_z with respect to σ_x and σ_y. For the three maps, the higher values are located along the boundaries of the considered area and on its corners. Observing Figure 4-39, it is worth noting the lower values of σ_z clearly evident for the pixels where the GCPs lie, comparing with respect to the surrounding pixels; this behaviour confirms the constraint function of the GCPs, which is even more evident along Z axis.

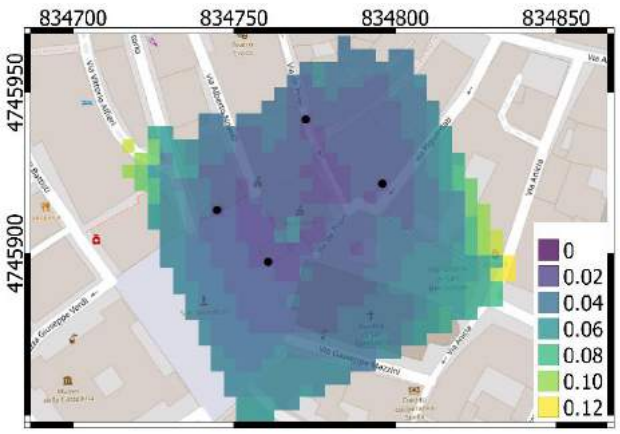


Figure 4-37 Estimated precisions [m] along X axis.

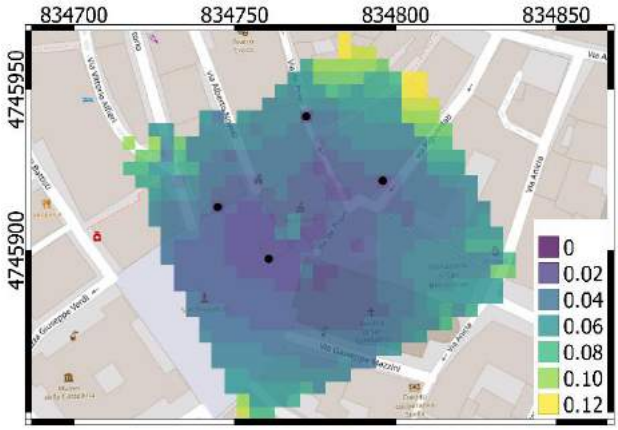


Figure 4-38 Estimated precisions [m] along Y axis.

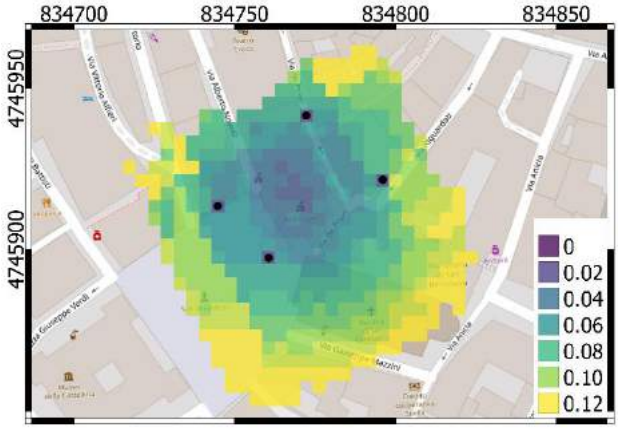


Figure 4-39 Estimated precisions [m] along Z axis.

In the latter case concerning San Salvatore Church, the visibility map suggests that the building is almost totally visible, thanks to the homogeneous heights of its ruins with respect to the ground, which are not likely to produce obstructions.

The values in the occurrence map (Figure 4-40) report the number of images that see each DSM cell, taking into account the presence of eventual occlusions. The black dots correspond to the resulting shooting point, obtained in the preliminary block design by U.Ph.O.

The average value of occurrence is 3.9, with a standard deviation of 2.0. The pattern of Figure 4-40 underlines what already mentioned regarding the Civic Tower, where the shooting geometry is similar to the present case: the cells with higher occurrence number are located in the centre of the survey area, whereas the occurrence number decreases moving towards the more external areas. For San Salvatore Church, the reduction of the number of occurrences

has a more gradual behaviour than the one of Civic Tower, also due to the different geometry and features of the two structures.

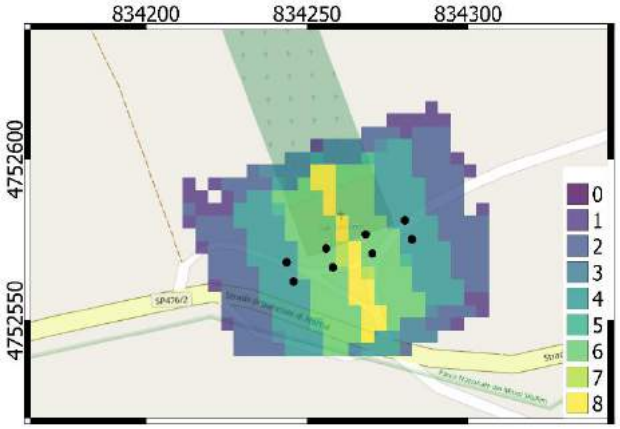


Figure 4-40 Occurrence map for San Salvatore Church.

The occurrence map has been classified, according to the classes reported in Table 4-7.

Table 4-7 Classification of occurrence map: cell count and percentage cover for the five classes.

Class	Occurrences nr.	Cell count	Percentage cover
0	0-1	27	7.8%
1	2	95	27.3%
2	3-4	116	33.3%
3	5-6	87	25.0%
4	7-8	23	6.6%

Less than 8% of pixels lie in class 0; the majority of the pixels falls in class 2, where 3 and 4 occurrences are contained, and a significant amount lie in class 3. None pixel is not seen (0 number of occurrences) and 20 pixels are seen by the highest number of images.

From Figure 4-41 to Figure 4-43, the estimated precision along X, Y, Z axes are shown, respectively. The average values of standard deviations are 2.1, 2.1 and 4.6 cm, respectively. As expected, higher values characterize σ_z with respect to σ_x and σ_y. For the three maps, the higher values are located along two opposite sides of the considered area. As highlighted in the previous test case, observing Figure 4-43, it is possible to notice the lower values of σ_z for the pixels where the GCPs lie.

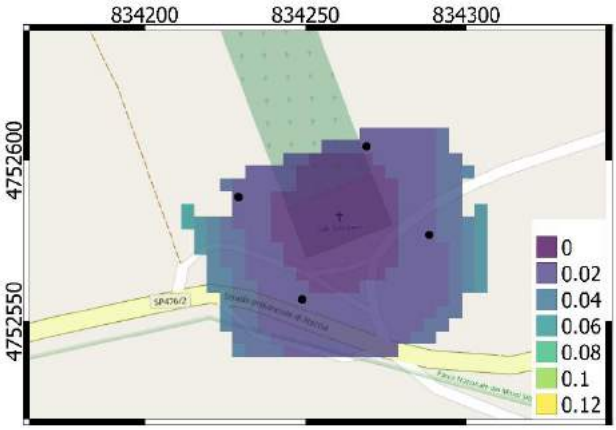


Figure 4-41 Estimated precisions [m] along X axis.

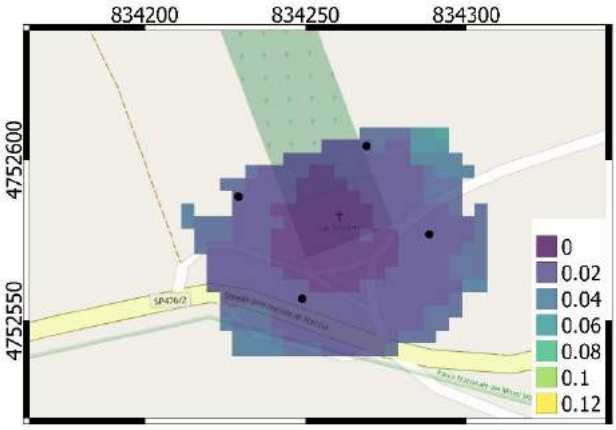


Figure 4-42 Estimated precisions [m] along Y axis.

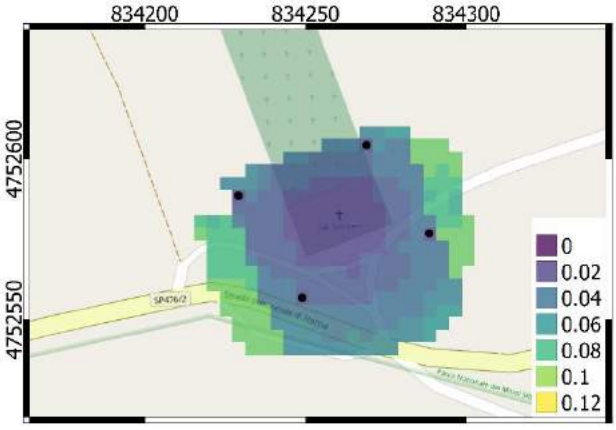


Figure 4-43 Estimated precisions [m] along Z axis.

In conclusion, the former case concerning Civic Tower more evidently highlights the consequences of the presence of obstructions, which highly influence the survey planning. In facts, the occurrence map and the precisions maps are affected by loss of coverage, precision and visibility, respectively.

On the contrary, San Salvatore Church, which is located in a flat and isolated area, is less impacted by obstructions, thanks to its configuration, which makes the survey planning easier and the achievable precisions higher than the previous case study.

4.2.2. Pompeii survey campaign

U.Ph.O. v.b has been integrated with the module for the management of vertical walls; it has been tested on the Pompeii case study, presented in paragraph 4.1.3, simulating the planning of the UAV survey campaign that was performed during the second visit on-site in September 2020.

UAV photogrammetry has been collected by means of a DJI Mavic 2 PRO drone, equipped with an embedded Hasselblad L1D-20c camera. The focal length is fixed at 10 mm. Four datasets have been acquired. In particular, the first is characterized by a nadir shooting geometry, in order to obtain a general framework of the area from 40 m above ground level (AGL), with a corresponding GSD of 10 mm. The other three datasets have been captured at a height of 15 m AGL, using both nadir and oblique shooting geometry. In the oblique configurations (Figure 4-44), a tilt of 45° has been applied and the drone path followed concentric trajectories looking towards first the inner and then the outer portion of analysed site. The obtained GSD is 4 mm, allowing a reliable site overview. The total amount of collected images is 1400.

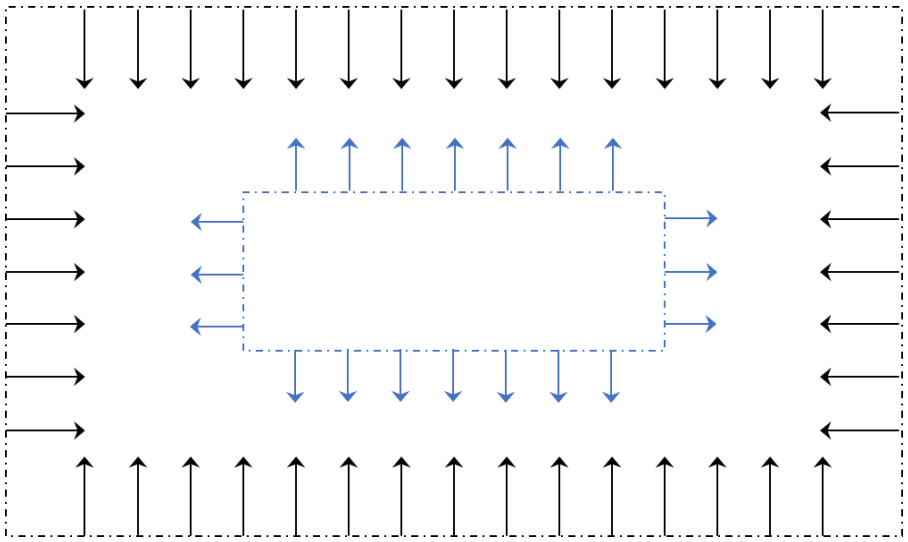


Figure 4-44. Drone paths (dashed lines) and images orientation (arrows) for the oblique configurations, towards the inner (black) and the outer (blue) directions.

The simulation of the survey planning is applied to three situations: the first concerns the general framework of the area, at a height of 40 m AGL, the second regards a more specific site overview at 15 m AGL; finally, the third is focused on a smaller sample, with a flight height of 15 m AGL. The chosen DSM resolutions are 5 m for the first two situations (Figure 4-45) and 0.5 m for the third scenario.

The camera parameters of Hasselblad L1D-20c have been given as input in U.Ph.O.. The overlapping parameters have been set to 80% and 60% for the longitudinal and transversal directions, respectively. A buffer along the boundaries of the area of interest has been considered, to avoid border effects in the study area limits.

The maps reported in the following have been produced by using QGIS, ver. 3.12 (QGIS Development Team, 2020), with the OpenStreetMap® (OpenStreetMap® Foundation, 2021) background. The coordinates are expressed in ETRF 2000-2008.0 reference frame, with UTM 32N projection.

Figure 4-45 DSM of the examined area at 5 m of resolution. Heights are included within the range 30.691-38.773 m.

As already stated, in the first and second scenarios the parameters are completely analogous, except for the flight height, set at 40 m and 15 m AGL, respectively. The third scenario, instead, is focused on a smaller sample (Figure 4-45), due to the huge computational effort required by passing from a DSM resolution of 5 m to 0.5 m (Figure 4-47). The location of the portion of interest with respect to the overall DSM is identified in Figure 4-46 with a red box. The other parameters have been maintained equal to the second scenario, thus a flight height of 15 m AGL has been considered.

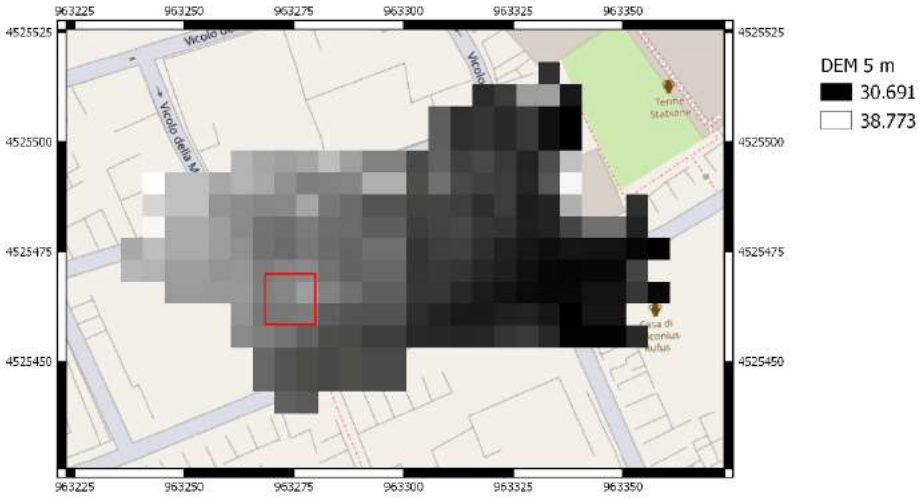


Figure 4-46 Location of the sample (red box) with respect to the overall DSM.

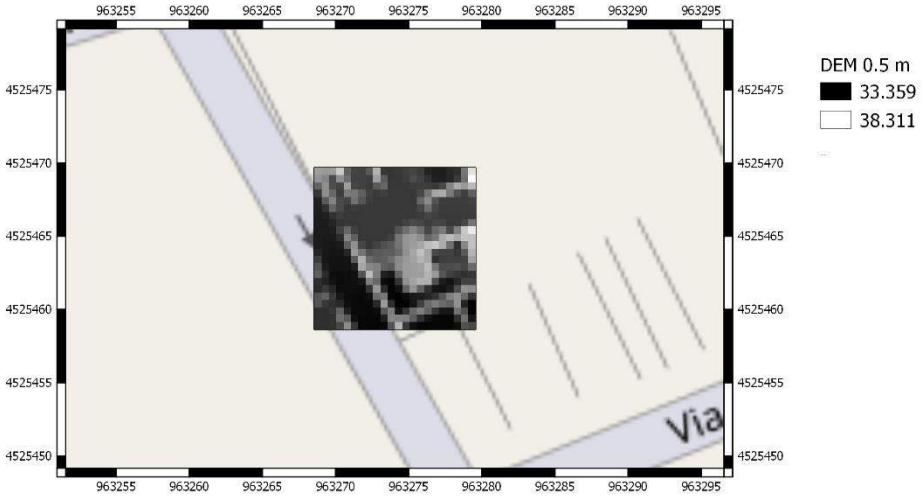


Figure 4-47 DSM of the sample at 0.5 m of resolution. Heights are included within the range 33.359-38.311 m.

All the case studies consider the nadir shooting geometry, organised in the typical photogrammetric block. Further developments will include the implementation of a module to allow the choice of a convergent geometry with respect to a specific point among the default options for the planning. The height parameter for the construction of voxels layers has been chosen coherently with the planimetric resolution, thus 5 and 0.5 m, respectively. The number of GCPs located within the study area corresponds to 11; all of them lie within the boundaries

of the original DSM, while just only one is included in the borders of the smaller sample depicted in Figure 4-47. Nevertheless, the presence of the whole group of GCPs was considered in both the configurations, accounting that the analysed area extension was the same for all the case studies, covering the dimension of the larger DSM (Figure 4-44). Thus, the mission planning considered the same drone path and, consequently, the same shooting centres for case studies nr.2 and nr.3, allowing to include in the least squares approach such images seeing the GCPs location. Undoubtedly, in case study nr.3 these outer portions with respect to the provided DSM need to be considered as not susceptible to obstructions because of the lack of information.

In Table 4-8, the summary of the employed parameters, common or specific for each case study, is reported.

Table 4-8 Input parameters for the three examined case studies.

Parameters	**Case nr.1**	**Case nr.2**	**Case nr.3**
Flight height [m]	40	15	
DSM resolution [m]	5		0.5
Camera	Hasselblad L1D-20c		
Overlapping rate	L: 80%; T: 60%		
Nr. GCPs	11		1 (+10)
Navigation GNSS accuracy [m]	10		

Deepening the first scenario with corresponding flight height of 40 m AGL and DSM resolution of 5 m, Figure 4-48 and Figure 4-49 show the number of occurrences, i.e., the number of images that see each cell, for the two obtained layers of voxels, together with the positions of the shooting centres. The related Table 4-9 contains the statistics about the cell count and percentage cover per each class of observed values.

It is worth noting that, in all the following figures, the layer 1 is the lowest, while the ones with increasing ID numbers are stacked above. Furthermore, the layer extension is determined starting from the minimum height value of the input DSM and adding the chosen resolution step-by-step, until the maximum height is included in the range of the top layer.

The colour scale of the legend is homogenized among all the experiments, to let a clearer and immediate comparison, whereas the related tables are case-dependent. Moreover, the percentage cover of the observed classes has been computed considering only the valid voxels for the current map, thus excluding no-data cells, which percentage, conversely, has been accounted for the total amount of analysed cells. No-data category includes such voxels whose maximum local height is lower than the minimum of the current layer.

Concerning the present case study, with flight height set at 40 m AGL, Figure 4-50 reports the differences between the values of occurrences in such voxels valid for both layers, i.e., the ones containing at least one data, to support further comments about the observed trend associated to the stacked layers. In this regard, it is worth noting that the values in Figure 4-50 are almost equal to Figure 4-49. In facts, the layer 1 is approximately entirely covered by the layer 2 stacking above. Figure 4-49 shows a huge area without data due insufficient heights of walls to be discretized in the second layer of voxels.

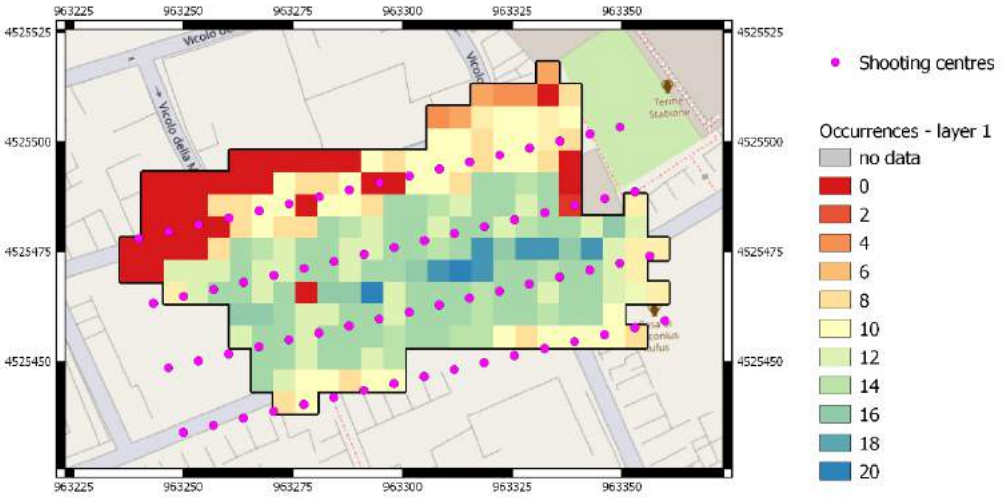


Figure 4-48 Occurrence map for the case study nr.1, layer 1.

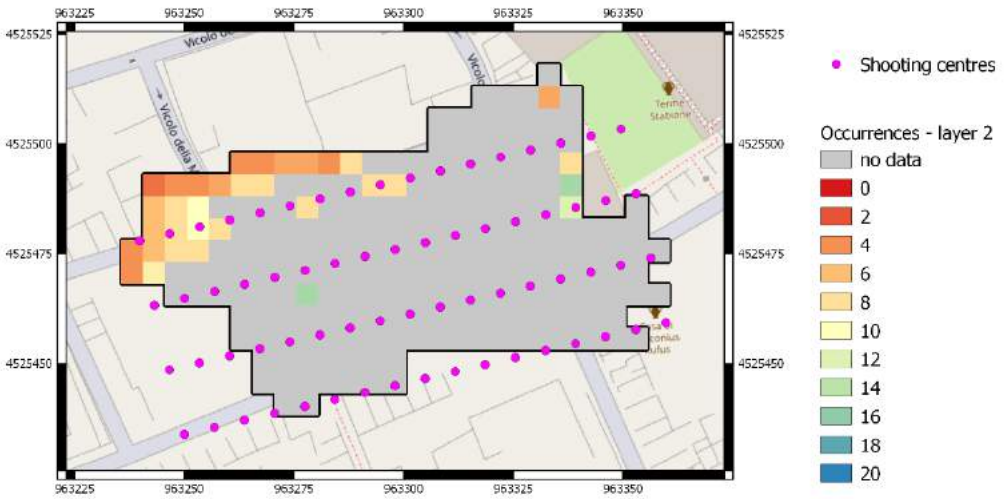


Figure 4-49 Occurrence map for the case study nr.1, layer 2.

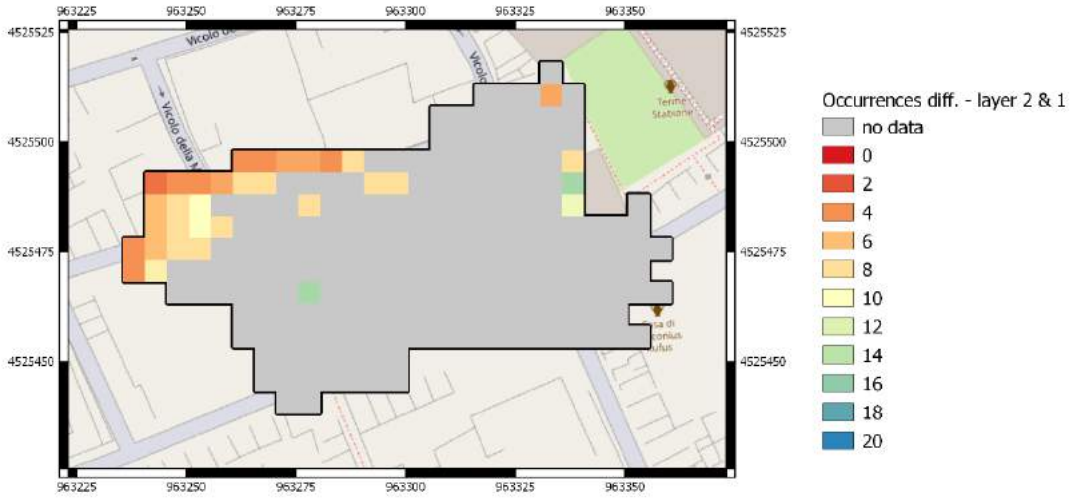


Figure 4-50 Resulting difference between layer 2 and 1 valid cells, case study nr. 1.

Table 4-9 Classification of occurrence map: cell count and percentage cover for the observed classes in case study nr.1.

Class	Nr. of occurrences	Layer 1		Layer 2	
		Cell count	% cover	Cell count	% cover
no data		0	0.0%	192	85.3%
0	0	32	14.2%	0	0.0%
1	1-2	1	0.4%	0	0.0%
2	3-4	3	1.3%	8	24.2%
3	5-6	3	1.3%	7	21.2%
4	7-8	20	8.9%	12	36.4%
5	9-10	42	18.7%	3	9.1%
6	11-12	32	14.2%	1	3.0%
7	13-14	14	6.2%	0	0.0%
8	15-16	70	31.1%	2	6.1%
9	17-18	0	0.0%	0	0.0%
10	19-20	8	3.6%	0	0.0%

In Table 4-9, it is possible noting that class 8 includes the higher percentage (31.1%, corresponding to 70 cells) of cells pertaining to layer 1, characterized by values within the range of 15-16 occurrences. Excluding such cells with no data, the most populated class for

layer 2 is class 4, containing values within the range of 7-8 occurrences (36.4%, corresponding to 12 cells). Moreover, in both Figure 4-48 and Figure 4-49 a high number of occurrences is shown even along the boundaries, typically around the value 10, while it is above 14 in the central block.

As already noticed, just a few walls stand out against the terrain above 4 m. In just one cell of the lowest layer the occurrence has been observed despite the presence of the stacking voxel above. The majority of the cells interested by this condition, thus developed onto two layers, is located in the north-west corner, coherently with the DSM trend (Figure 4-45). The terrain slope is ascending towards the west side, with an approximate average percentage of 4%. The interpretation of this result is that the block of adjacent cells pertaining to layer 2 does not allow framing the whole portion below, represented by the foot of these walls.

The analogous situation is examined in the second case study, setting the flight height at 15 m AGL. Observing Figure 4-51 and Figure 4-52, together with the related Table 4-10, it is evident that the number of shooting centres is substantially increased; nevertheless, the effect on the number of occurrences is not visible, due to the increased sensitivity to obstructions maintaining the same angle of view at a lower height. Moreover, since the UAV path elevation is considered as constant despite the DSM is not regular, with differences up to 8 m, the case study nr.2 is more affected by this condition, for its lower flight height. Thus, as consequence the real overlapping rate results significantly affected.

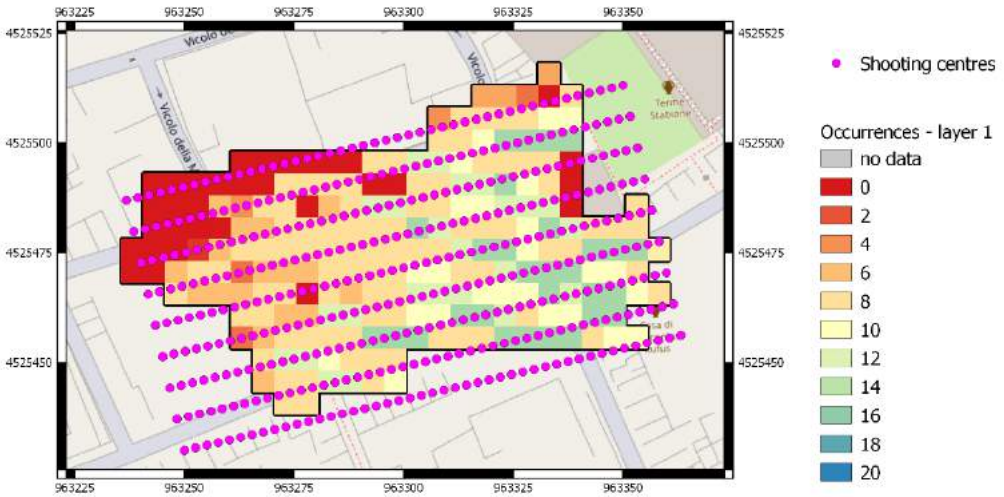


Figure 4-51 Occurrence map for the case study nr.2, layer 1.

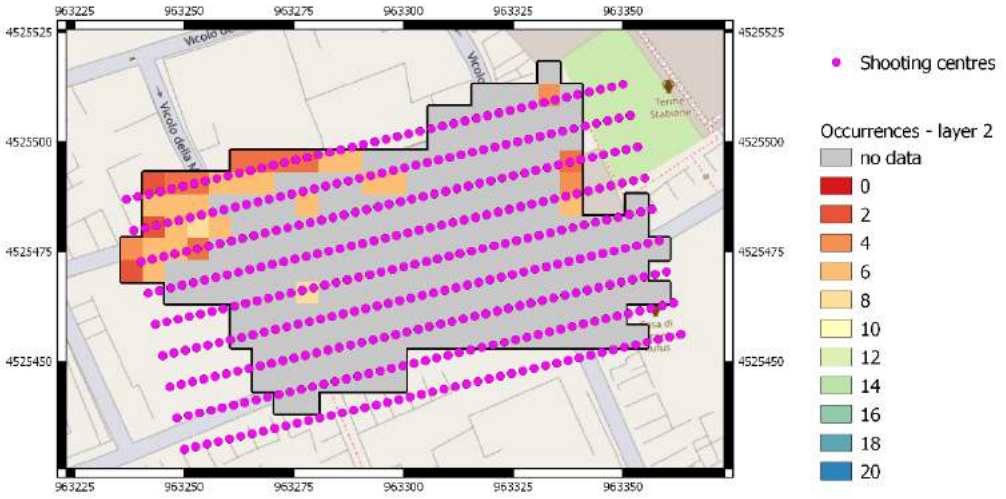


Figure 4-52 Occurrence map for the case study nr.2, layer 2.

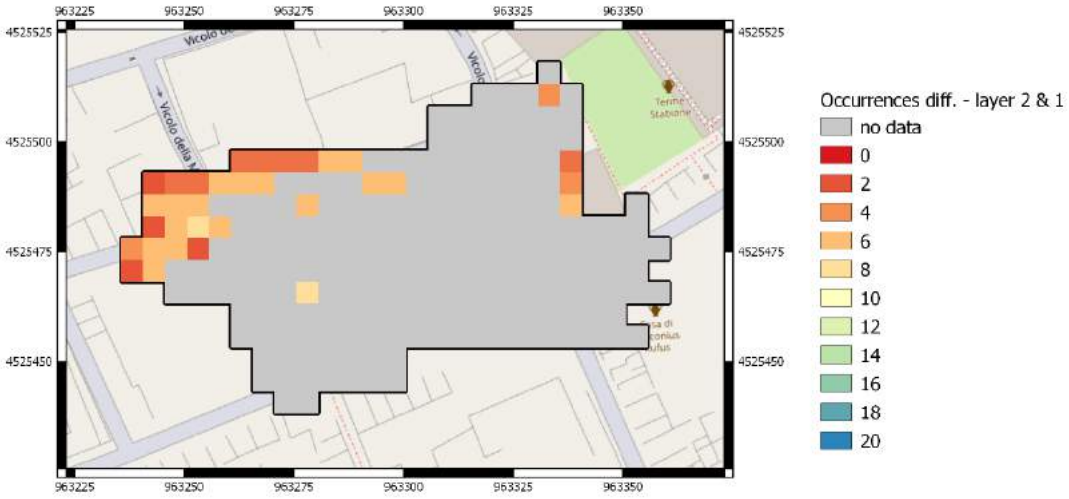


Figure 4-53 Resulting difference between layer 2 and 1 valid cells, case study nr. 2.

Table 4-10 Classification of occurrence map: cell count and percentage cover for the observed classes in case study nr.2.

Class	Nr. of occurrences	Layer 1		Layer 2	
		Cell count	% cover	Cell count	% cover
no data		0	0.0%	192	85.3%
0	0	32	14.2%	0	0.0%
1	1-2	1	0.4%	3	9.1%
2	3-4	5	2.2%	11	33.3%
3	5-6	20	8.9%	17	51.5%
4	7-8	88	39.1%	2	6.1%
5	9-10	26	11.6%	0	0.0%
6	11-12	24	10.7%	0	0.0%
7	13-14	0	0.0%	0	0.0%
8	15	29	12.9%	0	0.0%

Concerning the case study nr. 2, Figure 4-53 shows a trend similar to the one depicted in Figure 4-50. In fact, just one cell in the layer 1 is characterized by value 1, making the differences map approximately analogous to layer 2. In Table 4-10, the higher percentage (39.1%, corresponding to 88 cells) of cells pertaining to layer 1 is included in class 4, whose values belong to the range of 7-8 occurrences. Instead, excluding such cells with no data, the most populated class for layer 2 is class 3, containing values within the range of 5-6 occurrences (51.5%, corresponding to 17 cells). The maximum value is 15 occurrences, against the 20 of the previous case study, for the reasons explained before about the sensitivity to obstructions and the effective overlapping at the top of the structures.

The accuracy maps resulting from the first and the second case studies are reported in the following figures (from Figure 4-54 to Figure 4-59 for case study nr.1, from Figure 4-60 to Figure 4-65 for case study nr.2). Grey areas are no-data values which estimation has not been possible by means of least squares approach. Such cells with a single occurrence or no occurrences have been obviously not considered, together with the images having a single match. Starting from the second layer, grey cells include also such portions where maximum DSM height is lower than the interval of values included in the voxel level, as well as in the occurrence maps.

As a consequence of the lower number of occurrences per each cell, the second case study presents also a worse accuracy, as highlighted in the maps and in the resuming Table 4-11. Furthermore, in the west portion, characterized by height values in the order of 8 meters, the considered forward baseline and side wheelbase are not sufficient to reach the expected overlapping rate. In particular, $B = 2.640\,m$ and $i = 7.304\,m$ at a distance of 7 meters,

resulting from the difference of the flight height (15 m) and the DSM relative values from the take-off level (up to 8 m), let obtaining an effective overlapping in the order of 55% and 15% in longitudinal and transversal directions, respectively, just a few more than the minimum and, moreover, not considering the contribution of obstructions. Furthermore, in this compromised scenario the accuracy is very deteriorated due to the contribute of the term related to the baseline uncertainty that strongly affects σ, in a portion characterized by a low rate of occurrences pertaining to images almost isolated from the main part of the photogrammetric block, because of the obstructions produced by the surrounding high walls. Thus, the baseline uncertainty, deriving from the σ_x of the shooting centres coordinates set to 10 m, contributes to a significant increasing of the final σ because of the propagation of variance. In particular, starting from the equation 1.20, the uncertainty of B is much more significant if B, which is directly proportional to Z, is small.

In order to depict the values of the west area, the legend of the maps concerning case study nr.2 has been extended, since the previous scale limited at 0.05 m was not representative for this portion. The cells associated to a worse accuracy have been identified from red to magenta, up to 2 m.

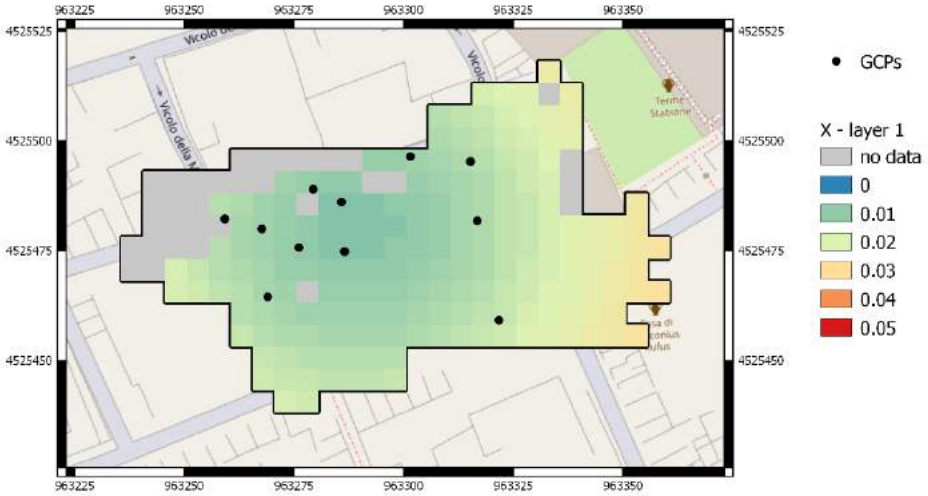


Figure 4-54 Accuracy [m] map for the X component of case study nr.1, layer 1.
Black dots represent GCPs, acquired by means of GPS in NRTK.

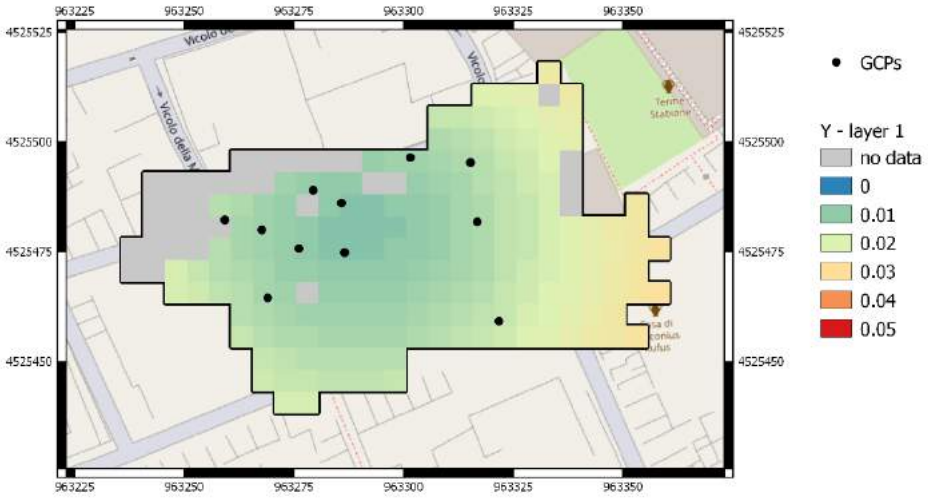


Figure 4-55 Accuracy [m] map for the Y component of case study nr.1, layer 1. Black dots represent GCPs, acquired by means of GPS in NRTK.

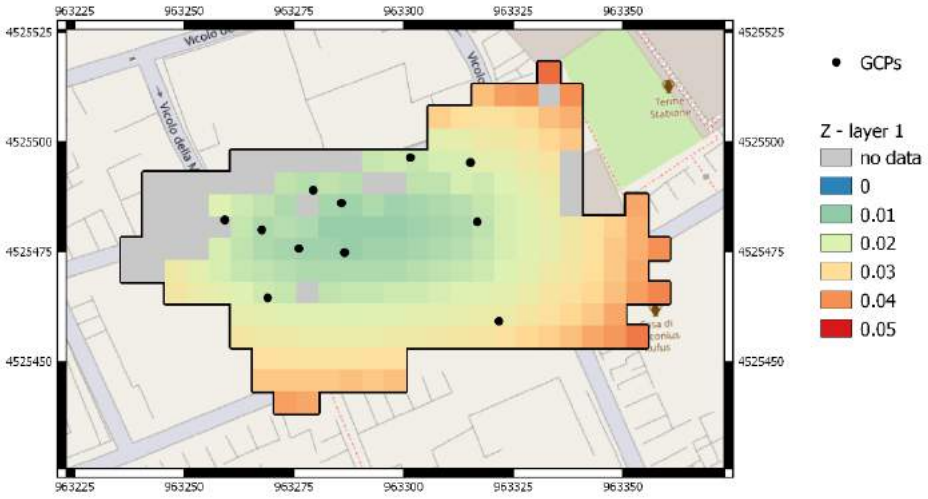


Figure 4-56 Accuracy [m] map for the Z component of case study nr.1, layer 1. Black dots represent GCPs, acquired by means of GPS in NRTK.

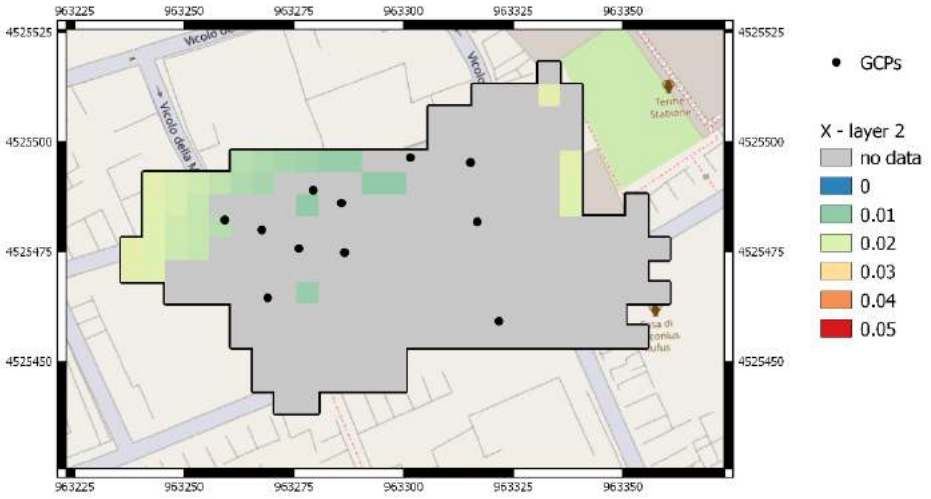


Figure 4-57 Accuracy [m] map for the X component of case study nr.1, layer 2. Black dots represent GCPs, acquired by means of GPS in NRTK.

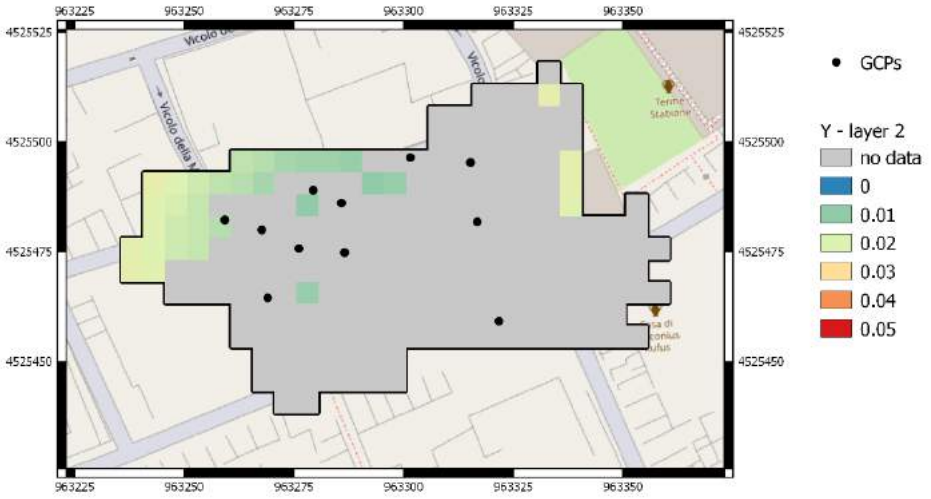


Figure 4-58 Accuracy [m] map for the Y component of case study nr.1, layer 2. Black dots represent GCPs, acquired by means of GPS in NRTK.

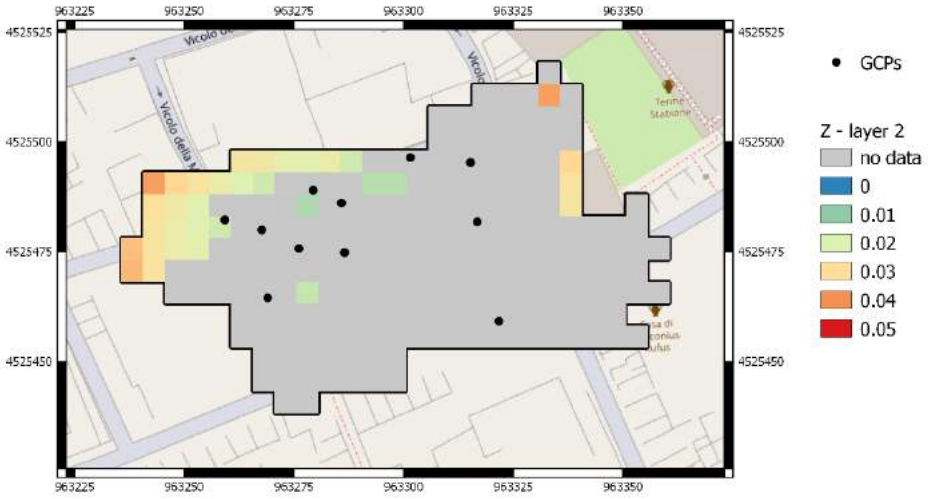


Figure 4-59 Accuracy [m] map for the Z component of case study nr.1, layer 2. Black dots represent GCPs, acquired by means of GPS in NRTK.

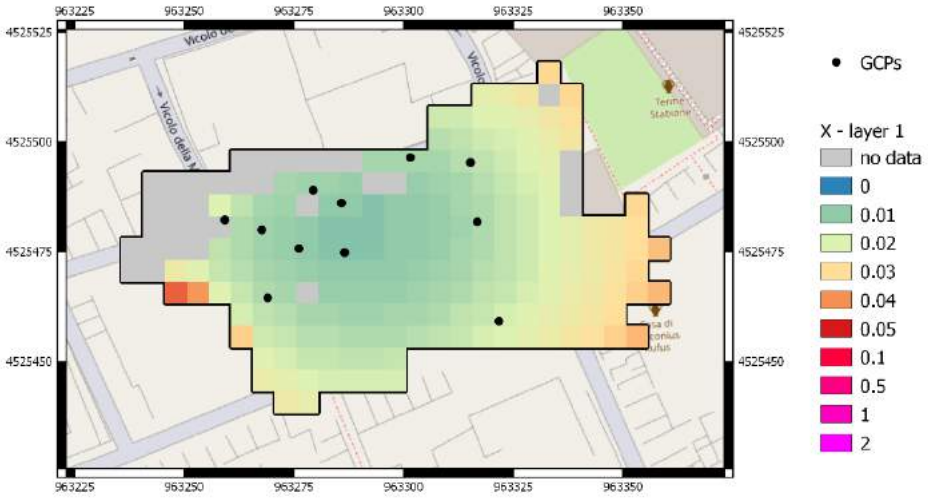


Figure 4-60 Accuracy [m] map for the X component of case study nr.2, layer 1. Black dots represent GCPs, acquired by means of GPS in NRTK.

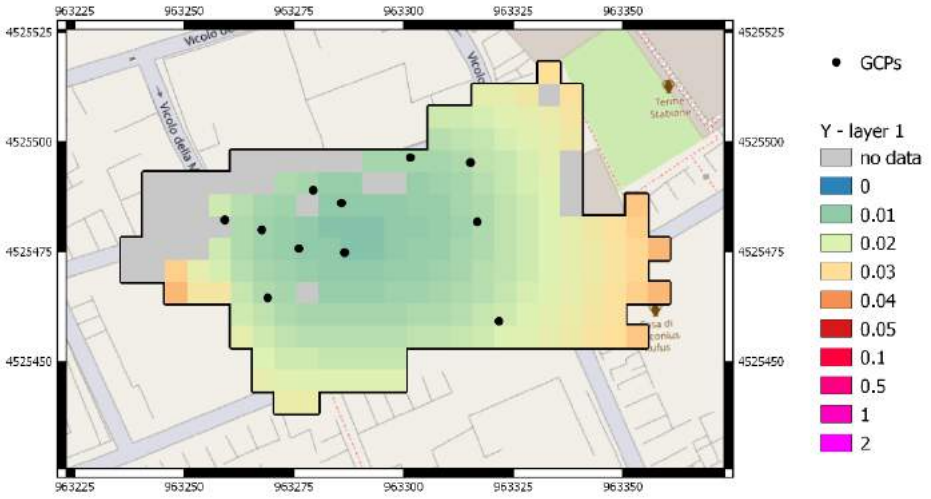


Figure 4-61 Accuracy [m] map for the Y component of case study nr.2, layer 1. Black dots represent GCPs, acquired by means of GPS in NRTK.

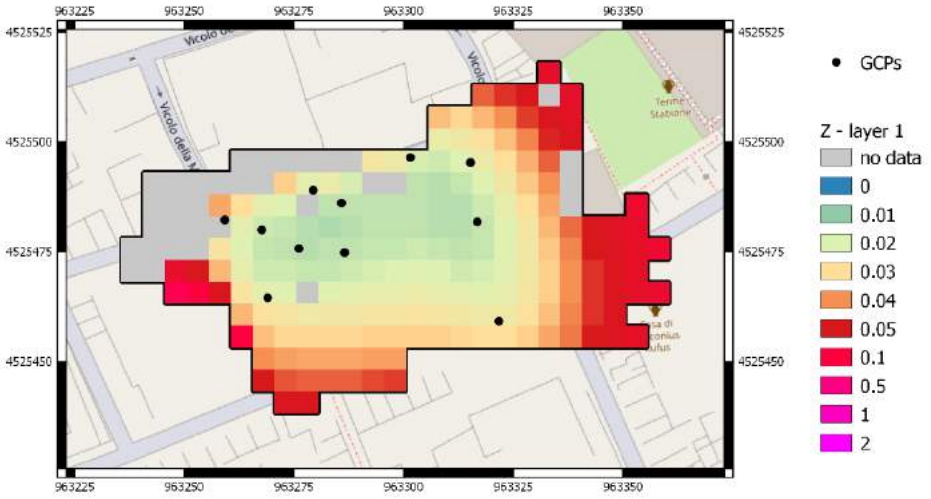


Figure 4-62 Accuracy [m] map for the Z component of case study nr.2, layer 1. Black dots represent GCPs, acquired by means of GPS in NRTK.

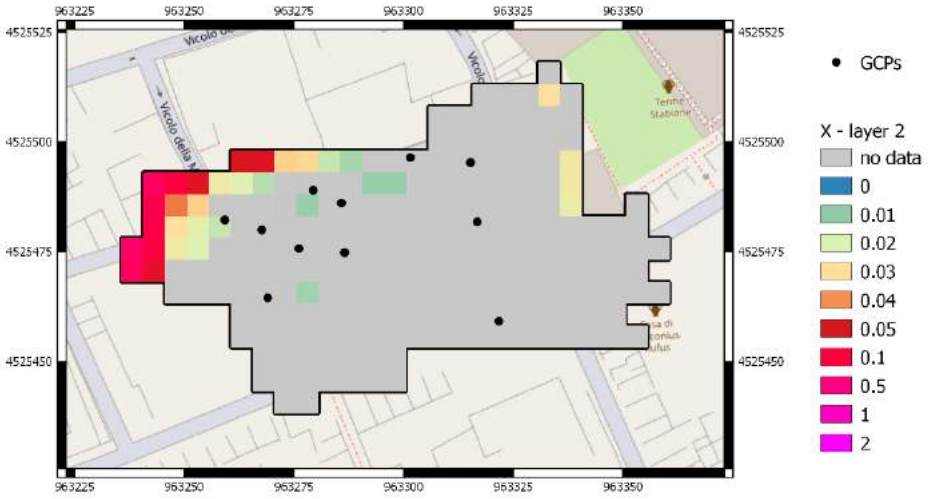


Figure 4-63 Accuracy [m] map for the X component of case study nr.2, layer 2. Black dots represent GCPs, acquired by means of GPS in NRTK.

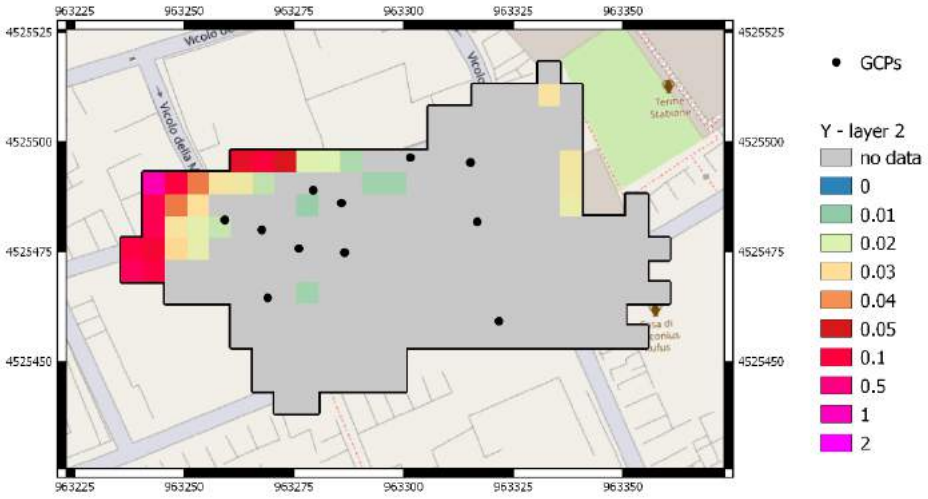


Figure 4-64 Accuracy [m] map for the Y component of case study nr.2, layer 2. Black dots represent GCPs, acquired by means of GPS in NRTK.

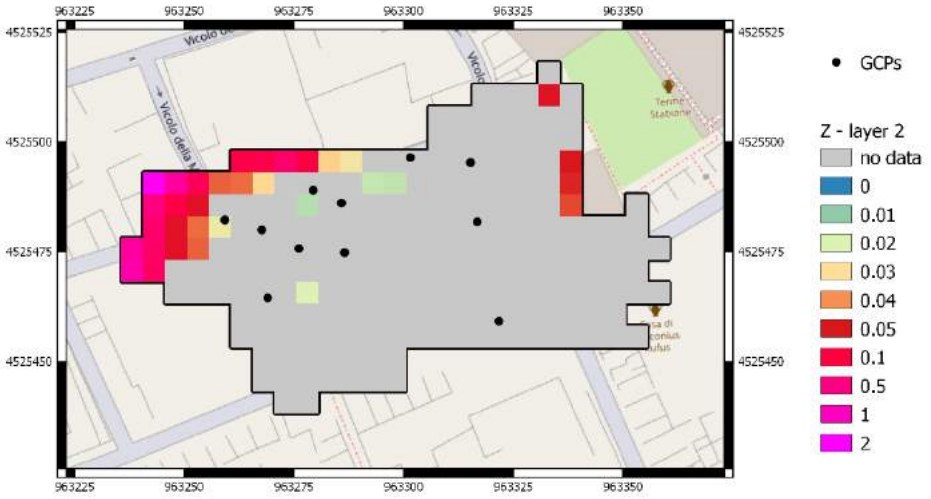


Figure 4-65 Accuracy [m] map for the Z component of case study nr.2, layer 2.
Black dots represent GCPs, acquired by means of GPS in NRTK.

Table 4-11 Accuracy statistics for each map of case studies nr.1 and 2.

	Case study 1		Case study 2	
	Layer 1	Layer 2	Layer 1	Layer 2
Stat.	Maximum values			
X [m]	0.029	0.023	0.044	0.357
Y [m]	0.029	0.023	0.035	0.892
Z [m]	0.043	0.038	0.235	1.810
Stat.	Minimum values			
X [m]	0.009	0.010	0.009	0.011
Y [m]	0.009	0.010	0.009	0.011
Z [m]	0.011	0.014	0.014	0.015
Stat.	Average values			
X [m]	0.017	0.016	0.020	0.055
Y [m]	0.017	0.016	0.020	0.060
Z [m]	0.024	0.024	0.043	0.180
Stat.	Standard deviation			
X [m]	0.001	0.000	0.003	0.025
Y [m]	0.001	0.000	0.001	0.086
Z [m]	0.003	0.002	0.024	0.173

In the maps related to layer 1 (Figure 4-54Figure 4-56 and Figure 4-60Figure 4-62), it is possible noting that the central portion of the examined area is characterized by values below 1-3 cm, according to the examined component and case study. The trend worsens approaching the boundaries, where the maximum values are located.

Concerning layer 2, despite the maps (Figure 4-57Figure 4-59 and Figure 4-63Figure 4-65) present many groups of pixels mutually disconnected, a similar trend is visible, with the maximum values located close to the global boundaries.

The case study nr.1 shows a very similar trend between layer 1 and 2, while in case study nr.2 the layer 2 statistics result strongly affected by the presence of cells which accuracy is substantially depleted because of the overlapping not fitting to the theoretical value. The most evident deterioration concerns the Z component, in which the maximum value reach 1.81 m, leading to an average of 0.18 m and a standard deviation of 0.17 m. Also, the Y component is significantly worse than X, being almost parallel to the transversal and longitudinal flight direction, respectively. In these cases, the maxima reach 0.89 and 0.36 m and the standard deviation 0.09 and 0.03 m, respectively, while the average is 0.06 for both cases.

Considering all the exposed maps and tables, it is worth noting that the statistic and graphic trends are in favour of the first solution, despite it contemplates a higher flight height, resulting in a worse resolution on ground and in a larger GSD. In conclusion, in the proposed test the contribution of the reduction of the relative flight height, which bring to a lower standard deviation on the observables, it is not sufficient to counteract the decreased occurrence on the single voxels, due to the raising of sensitivity to obstructions. Anyway, the present work is effectively a network simulation, thus it performs the analysis on the basis of the inverse of the normal matrix of the least squares approach, which is representative of the covariance matrix of the estimations of unknown parameters (C_{xx}) unless the estimation of σ_0^2 a-posteriori.

As known, the estimation of C_{xx} is not affected by the particular choice of σ_0^2 a-priori, which represents the quality of the observables, contrarily to σ_0^2 a-posteriori; thus, the coherent comparison of the inverse of the normal for the two scenarios needs to consider the variation of σ_0^2 a-posteriori. Specifically, the inverse of the normal for the two case studies are representative of the C_{xx} only assuming that the σ_0^2 a-posteriori are comparable. Its estimation is mainly affected by the ratio between the magnitude of the single discards (v) related to the equations of observation and their numerousness (numerator) and the redundancy (denominator). Thus, it could be considered equivalent to:

$$\sigma_0^2 = \frac{v^2 \cdot (\text{occurrences} \cdot 2)}{(\text{occurrences} \cdot 2) - \text{unknowns}} \tag{4.2}$$

where the occurrences are multiplied by two in order to include in the observables both ξ and η image coordinates, while the unknowns consist of the three object coordinates for each ground point and six EO parameters for each image.

While the former component, ascribable to the object coordinates, is the same for the two case studies, considering as input the same DSM at the same resolution, conversely the number of images is substantially increased in the second scenario, however against a reduced number of occurrences. Therefore, the ratio between these parameters depends on the detailed evaluation of the realistic obstructions, as in the purpose of U.Ph.O. tool.

In the following Table 4-12, the values obtained for the analysed case studies are reported.

Table 4-12 Parameters of least square approach of case studies nr.1 and 2.

	Case study 1	Case study 2
Observables (excluding pseudo-observations)	3372	2614
Unknowns	834	1710
Redundancy	2538	904

From Table 4-12, it is possible to infer that, hypothesizing a comparable magnitude of the discards *v* for the two scenarios, the redundancy, i.e., the denominator, is higher for the first case study, thus it produces the decrease of σ_0^2 a-posteriori. The illustrated steps allow to obtain an indicative evaluation of the C_{xx} matrix in both the case studies, starting from the respective inverses of normal matrices.

Further investigations about the effective comparability of the discards will be faced in future by enhancing the check module, already included in U.Ph.O. procedure, for their realistic evaluation in light of the σ_0^2 a-posteriori estimation. Basing on the proposed observations, it could be asserted that the usefulness of U.Ph.O. for mission planning is particularly relevant due to the non-triviality of the choice of optimal parameters for the survey campaign.

Since the cell of 5 m is not optimal to highlight the presence of vertical walls, but rather to examine the overall trend of the area, a further case study (nr.3) has been considered with a DSM resolution of 0.5 m, which results more suitable to understand the potentialities of the extra module. Since the number of voxel layers is substantially increased due to the higher resolution, despite the global maximum is higher for this restricted portion, just two layers are taken in consideration as reference for the present analysis. The complete list of images related to all the stacked layers is reported in the dedicated Appendix.

Figure 4-66 and Figure 4-67 depict the occurrences maps for the layers 1 and 2, together with the position of the shooting centres (magenta dots). Their resulting difference is reported

in Figure 4-68, while statistics about the cell count and the percentage cover are listed in Table 4-13. As specified for the previous Figures, the layer 1 is the lowest, while the increasing numeration identifies the stacked voxels.

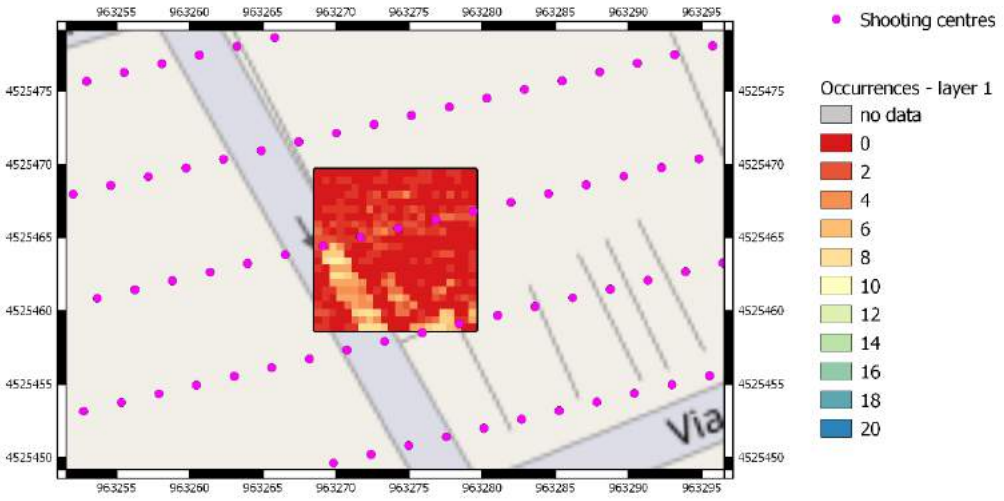


Figure 4-66 Occurrence map for the case study nr.3, layer 1.

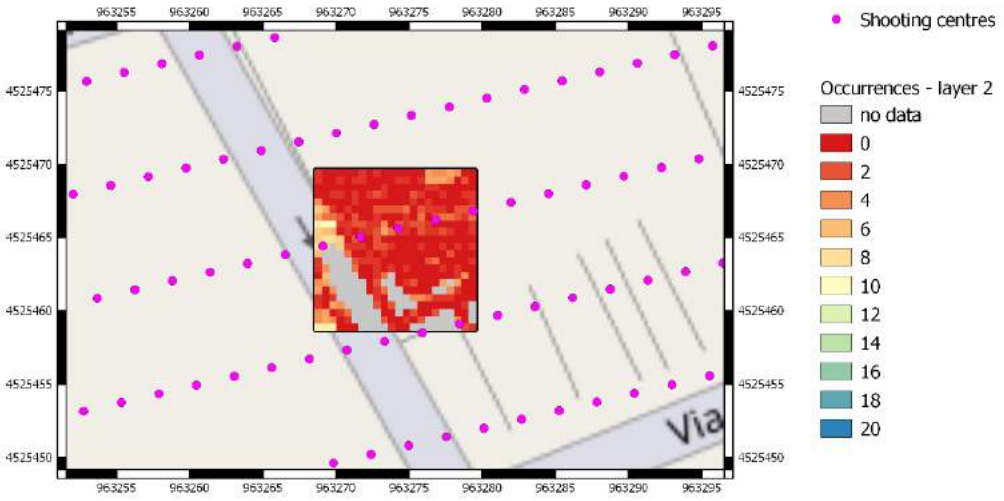


Figure 4-67 Occurrence map for the case study nr.3, layer 2.

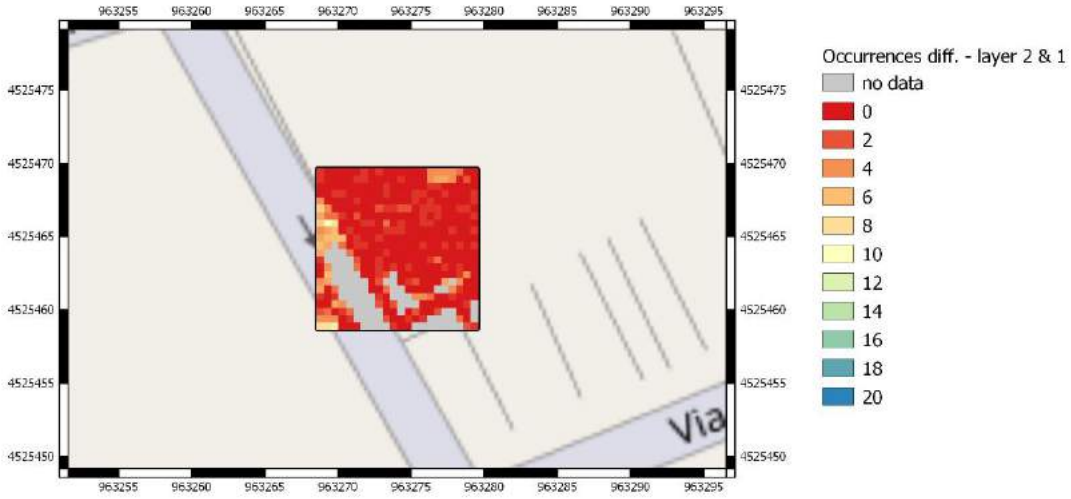


Figure 4-68 Resulting difference between layer 2 and 1 valid cells, case study nr. 3.

Table 4-13 Classification of occurrence map: cell count and percentage cover for the observed classes in case study nr.3, layers 1 and 2.

Class	Nr. of occurrences	Layer 1		Layer 2	
		Cell count	% cover	Cell count	% cover
no data		0	0.0%	76	15.7%
0	0	261	53.9%	213	52.2%
1	1-2	140	28.9%	128	31.4%
2	3-4	35	7.2%	33	8.1%
3	5-6	23	4.8%	19	4.7%
4	7-8	23	4.8%	9	2.2%
5	9-10	2	0.4%	6	1.5%

It is evident from Table 4-13 that the majority of the valid cells of both layers (>50%) are obscured by the ones stacking above. Nevertheless, the global number of occurrences is up to 10. The different resolution allows a better evaluation of the effective hiding due to the development of high walls. The phenomenon of depleted accuracy noticed in case study nr.2 is not observed in this configuration, thus even if the design overlapping is not always maintained, the obtained accuracy is always equal or below 5 cm for all the analysed layers.

Figure 4-68 depicts the difference between the first two stacking layers, highlighting that, conversely than the cases with 5 m of DSM resolution, in this scenario a certain number of cells is visible despite the presence of the stacking layer.

Table 4-13 allows affirming that the voxels have occurrences included within the range from 0 to 10 for both the first and the second layers, with a denser concentration at lower classes. Referring to the analogous Tables in the Appendix, related to the occurrences trend in the upper layers, the percentage cover of no-data values is obviously increasing as a consequence of the local maximum that has been already reached in the lower layers.

The accuracy values along the three axes X, Y, and Z for case study nr.3 are represented from Figure 4-69 to Figure 4-74. Table 4-14 shows the statistics, including maximum, minimum, average and standard deviation related to each map.

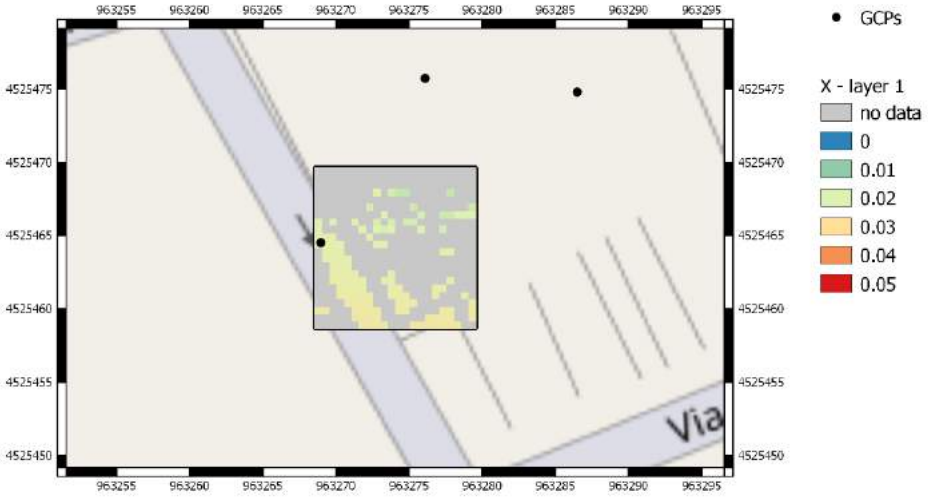


Figure 4-69 Accuracy [m] map for the X component of case study nr.3, layer 1. Black dots represent GCPs, acquired by means of GPS in NRTK.

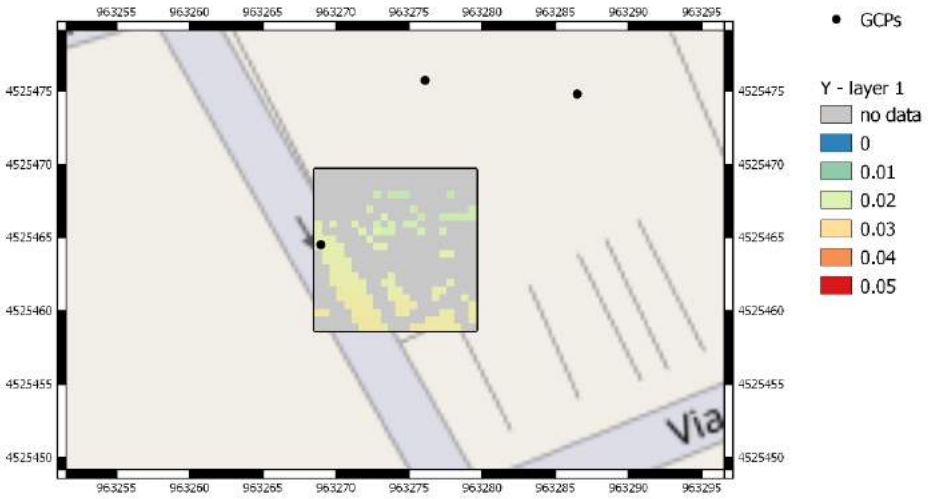


Figure 4-70 Accuracy [m] map for the Y component of case study nr.3, layer 1. Black dots represent GCPs, acquired by means of GPS in NRTK.

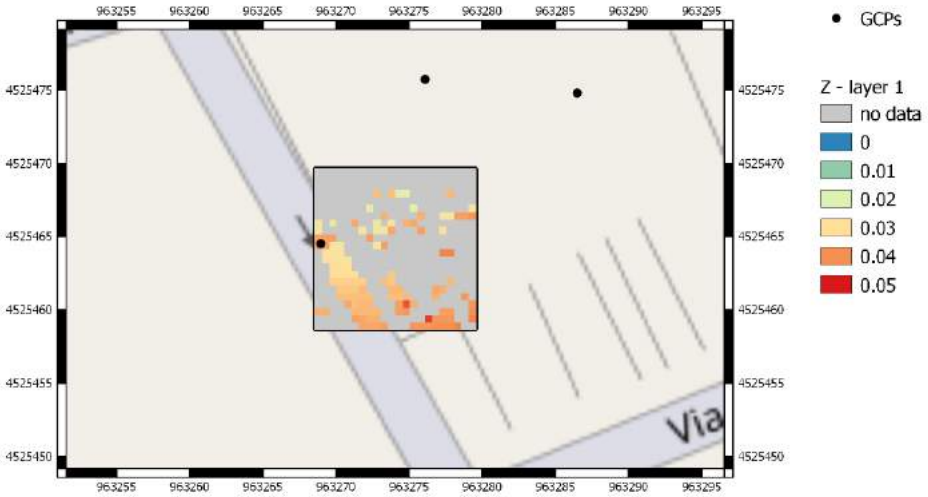


Figure 4-71 Accuracy [m] map for the Z component of case study nr.3, layer 1. Black dots represent GCPs, acquired by means of GPS in NRTK.

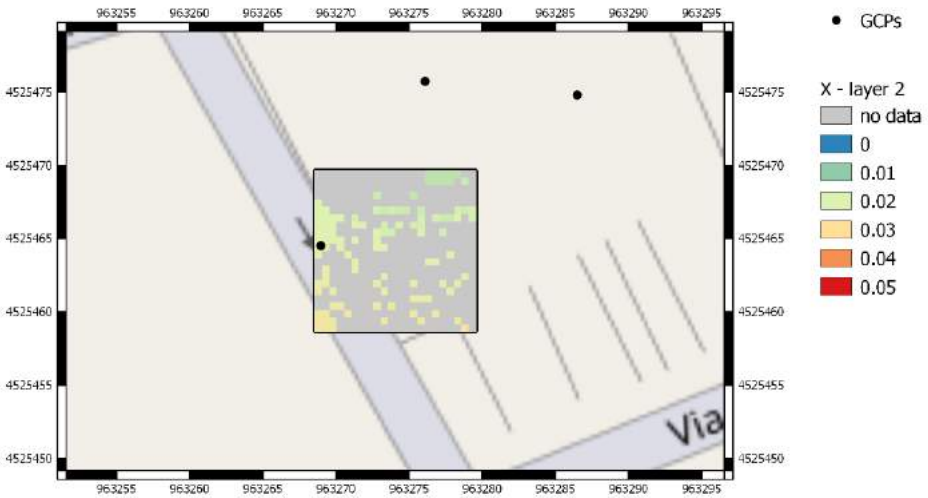


Figure 4-72 Accuracy [m] map for the X component of case study nr.3, layer 2. Black dots represent GCPs, acquired by means of GPS in NRTK.

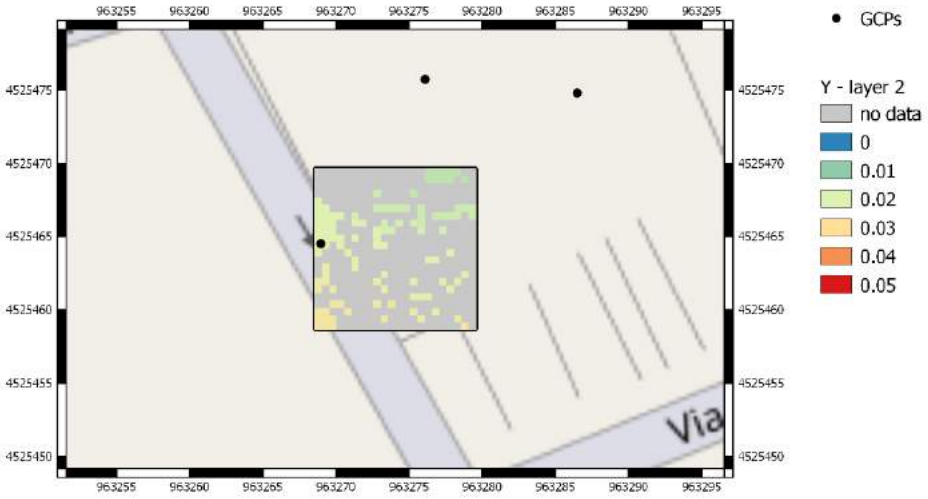


Figure 4-73 Accuracy [m] map for the Y component of case study nr.3, layer 2. Black dots represent GCPs, acquired by means of GPS in NRTK.

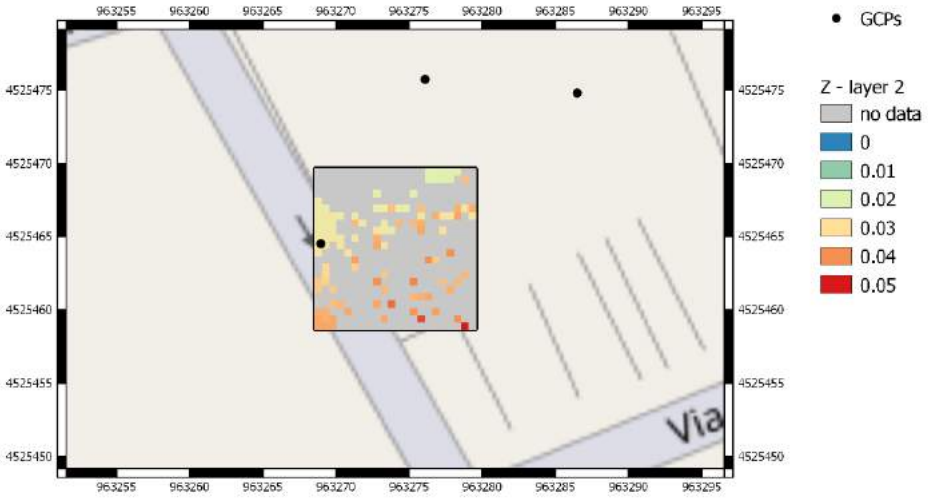


Figure 4-74 Accuracy [m] map for the Z component of case study nr.3, layer 2. Black dots represent GCPs, acquired by means of GPS in NRTK.

Table 4-14 Accuracy statistics for each map of case study nr.3.

	Case study 3				
	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
Stat.	Maximum values				
X [m]	0.026	0.028	0.027	0.027	0.026
Y [m]	0.027	0.028	0.027	0.027	0.026
Z [m]	0.047	0.051	0.042	0.041	0.040
	Layer 6	Layer 7	Layer 8	Layer 9	Layer 10
Stat.	Maximum values				
X [m]	0.027	0.027	0.027	0.018	0.022
Y [m]	0.027	0.026	0.026	0.019	0.021
Z [m]	0.039	0.039	0.039	0.024	0.029
	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
Stat.	Minimum values				
X [m]	0.017	0.016	0.016	0.017	0.017
Y [m]	0.017	0.016	0.016	0.017	0.017
Z [m]	0.021	0.020	0.020	0.021	0.021
	Layer 6	Layer 7	Layer 8	Layer 9	Layer 10
Stat.	Minimum values				
X [m]	0.017	0.018	0.019	0.018	0.019
Y [m]	0.018	0.018	0.019	0.019	0.020
Z [m]	0.021	0.020	0.021	0.024	0.024
	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
Stat.	Average values				
X [m]	0.023	0.021	0.020	0.022	0.022
Y [m]	0.023	0.021	0.020	0.021	0.021
Z [m]	0.035	0.031	0.027	0.031	0.030
	Layer 6	Layer 7	Layer 8	Layer 9	Layer 10
Stat.	Average values				
X [m]	0.022	0.021	0.022	0.018	0.020
Y [m]	0.021	0.021	0.022	0.019	0.021
Z [m]	0.030	0.028	0.030	0.024	0.027

	Case study 3				
	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
Stat.	Standard deviation				
X [m]	0.001	0.001	0.001	0.001	0.002
Y [m]	0.001	0.001	0.001	0.002	0.001
Z [m]	0.002	0.003	0.002	0.003	0.003
	Layer 6	Layer 7	Layer 8	Layer 9	Layer 10
Stat.	Standard deviation				
X [m]	0.002	0.001	0.000	0.000	0.000
Y [m]	0.001	0.001	0.000	0.000	0.000
Z [m]	0.003	0.002	0.001	0.000	0.000

It is possible noting that the global maximum values of the maps pertaining to case study nr.3 are almost coherent with the ones observed in case study nr.1. The Z accuracy values are just slightly worse than the planimetric precisions. Moreover, the standard deviation is equal or lower than 3 mm for all the maps. The average value is up to 35 mm for all the layers and components, while the maximum values are within the range 17–23 mm for X and Y and 20–35 mm for Z.

Finally, a further group of images has been created in order to depict the analogous values proposed in accuracy maps related to case study nr.3, but from a different perspective. In particular, a portion of wall, enclosed in the red box in Figure 4-75, has been analysed looking at the YZ plane, instead of the stacking layers parallel to XY considered for the previous maps. This is a further enhancement in order to have an improved readability of the map by representing it along the direction of the examined façade. Figure 4-76, Figure 4-77 and Figure 4-78 depict the X, Y, and Z components data, respectively. The extraction of these maps has been achieved automatically in U.Ph.O. workflow.

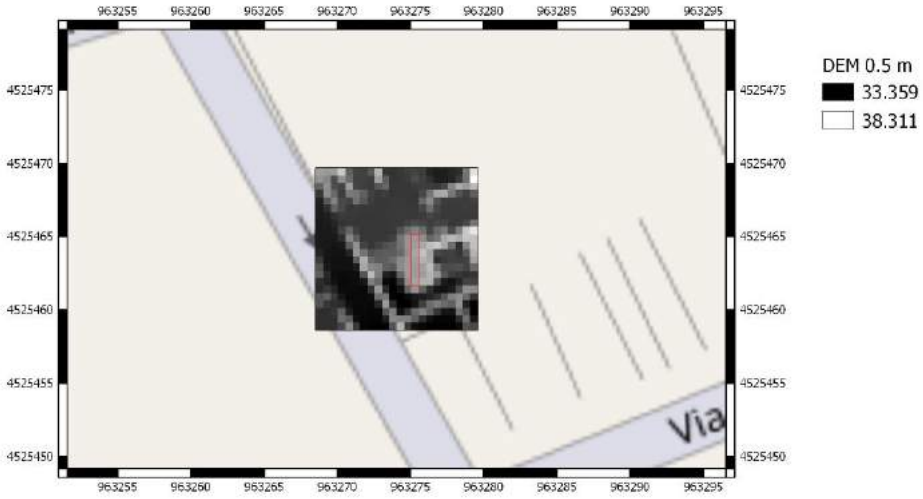


Figure 4-75 Highlighting of the portion of interest for YZ maps of accuracy (red box).

Figure 4-76. Accuracy [m] map for the X component of case study nr.3, parallel to YZ plane of the façade of interest.

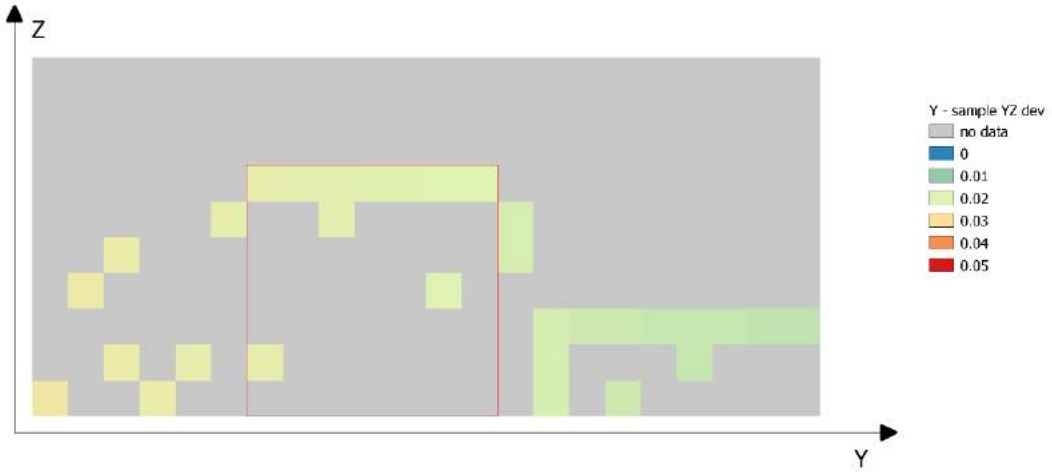


Figure 4-77 Accuracy [m] map for the Y component of case study nr.3, parallel to YZ plane of the façade of interest.

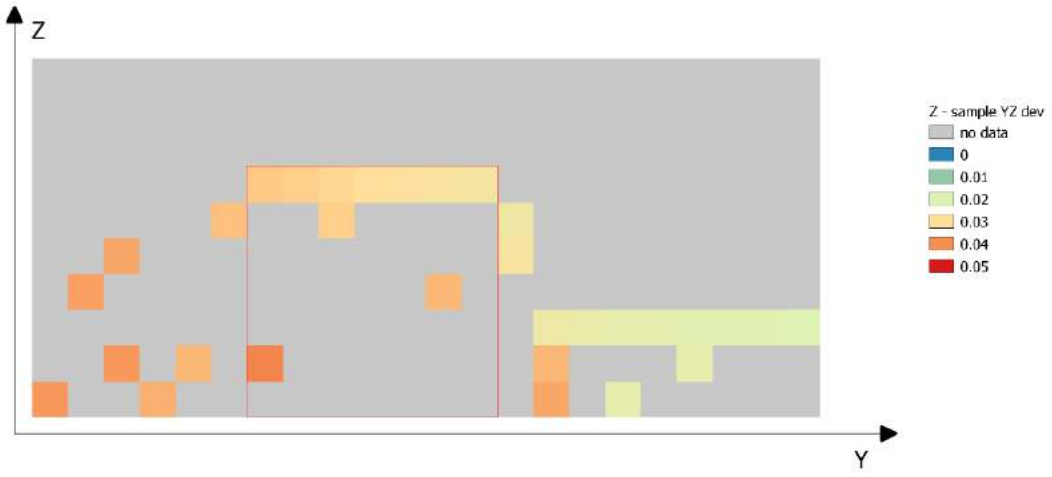


Figure 4-78 Accuracy [m] map for the Z component of case study nr.3, parallel to YZ plane of the façade of interest.

CONCLUSIONS AND FUTURE PERSPECTIVES

The present work describes a suite of tools focusing on the whole photogrammetric workflow, with particular attention to the UAV survey, from the planning to the final restitution. The final aim is to give to a medium-skilled user the instruments to integrate the available techniques by freely customizing the pre and post processing procedures, making them more efficient. The main strategies of the software, concerning all the different phases of planning, acquisition, post-processing, and restitution, join the fundamental principles of Photogrammetry with Computer Vision approach, adopting machine-learning processes to iteratively adjust the parameters evaluation on the level of detail of the solution.

The three tools composing the suite are MAGO (Adaptive Mesh for Orthophoto Generation), U.Ph.O. (Unmanned Photogrammetric Office), mostly implemented by Eng. Daniele Passoni, and a third work-in-progress procedure to achieve real-time positioning and modelling by means of a Bundle Block Adjustment approach. Despite the variety of the treated topics, there is a common thread characterizing the implementation of the tools and consisting in two basic strategies, founded on the reconstruction of an adaptive mesh and on the concepts of rotation convention and orthogonal projection.

In particular, the tool MAGO has been fully developed during the PhD course, relying on the support of existing and well-established libraries and sources. In facts, the code, written in C++ language and consisting in about 3500 lines, has been implemented by using the open source library OpenCV. The peculiarity of this tool, compared to the multitude of existing ones, is the possibility to automatically obtain an “unrolled” joint orthophoto of adjacent façades. The input source is a point cloud, thus, the procedure allow avoiding approximations typically attributable to the projection of images on the ready-made polygonal mesh. In facts, a step-by-step adaptive mesh is reconstructed by using the three best-fitting points according to specific criteria and thresholds.

The strategies of adaptive mesh and ortho-projection will be further deepen in the production of the Bundle Block Adjustment based tools, dedicated to real-time navigation and modelling, respectively. These instruments, which implementation and test are still at the beginning due to the hindrance of Covid-19 pandemic outbreak, are intended to exploit the Bundle Block Adjustment approach in two mirrored perspectives, named active and passive according to the movement of involved sensors, i.e., cameras. On the one hand, the navigation system itself is moving accordingly with the object, playing an “active” role, and providing instantaneously the positioning; on the other hand, the acquisition of images for the real-time 3D modelling of an evolving scene considers the use of stationary cameras.

The last goal of the present work has been the enrichment of an already existent tool, U.Ph.O. (Unmanned Photogrammetric Office), implemented within his PhD thesis by Eng. Daniele Passoni in Matlab environment, for planning and checking the survey campaign

accuracy. Despite the existence of a huge number of tools dedicated to mission planning, the distinguishing feature of U.Ph.O. is the particular focus on the aspect of expected precision, which estimation is based on a rigorous network simulation taking into account the peculiarity of the survey conditions. The usefulness of this approach is justified by the complex interaction of the several involved parameters, which reciprocal influence is far from trivial.

The contribution to this procedure has been represented by an adding module to obtain the accuracy evaluation along the vertical walls. The tool requires as input the DSM of the area of interest. In the basic version, the heights are taken into account to automatically obtain the realistic obstruction for each object point with respect to images, for the least squares simulation. In the new module the implemented strategy permits to evaluate the accuracy at different heights of the wall, hypothesizing the presence of lower levels under the maximum height of each cell, until the global minimum.

The experimental tests have been satisfying, as widely shown by means of graphics and statistics about the reached accuracy and quality.

In the present thesis work, in particular, the logical and chronological evolution of the tools has been highlighted, as well as the reciprocal common thread among the treated topics.

The future perspectives concern the improvement of such aspects that have been still pointed out as criticalities, but especially the sharing of this suite of tools to the scientific community. The feedback coming from the geomatics experts would be a fundamental step to validate the tools, proving their applicability to generalized case studies.

REFERENCES

Ackermann, F., Ebner, H., Klein, H., 1973. Block triangulation with Independent Models.

Agisoft© LLC, 2019. Agisoft Metashape Professional Software, Version 1.5.1. agisoft.com (1 March 2021).

Ajmar, A., Arco, E., Boccardo, P., Giulio Tonolo, F., Yoong, J., 2019. Updating a road network dataset exploiting the results of semantic segmentation techniques applied to street-level imagery. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W13, 1511–1516. doi.org/10.5194/isprs-archives-XLII-2-W13-1511-2019

Alberda, J. E., 1980. A review of analysis techniques for Engineering control schemes, Proceedings of the Industrial and Engineering Survey Conference, London.

Alidoost, F., Arefi, H., 2017. Comparison of UAS-based Photogrammetry software for 3D point cloud generation: a survey over a historical site, ISPRF Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-4/W4, 55-61. 10.5194/isprs-annals-IV-4-W4-55-2017

Alsadik, B., Karam, S., 2021. The Simultaneous Localization and Mapping (SLAM)-An Overview. Surveying and Geospatial Engineering Journal, 2(01), 01 - 12. https://doi.org/10.38094/sgej1027

Amhar F., Jansa J., Ries C., 1998. The generation of true orthophotos using a 3D building model in conjunction with a conventional DTM. International Archives of Photogrammetry and Remote Sensing, 32(4), pp. 16-22

Ao, W., Wang, L., Shan, J., 2019. Point cloud classification by fusing supervoxel segmentation with multi-scale features. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W13, 919–925. doi.org/10.5194/isprs-archives-XLII-2-W13-919-2019

Azhari, F., Kiely, S., Sennersten, C., Lindley, C., Matuszak, M., Hogwood, S., 2017. A comparison of sensors for underground void mapping by unmanned aerial vehicles. In Underground Mining Technology 2017 – Hudyma, M., Potvin, Y., © 2017 Australian Centre for Geomechanics, Perth, ISBN 978-0-9924810-7-0

Barazzetti L., Brovelli M., Scaioni M., 2007. Problems Related to the Generation of True-Orthophotos with LiDAR DDSMs. In: Proc. ISPRS Work. "Laser Scanning 2007 and SilviLaser 2007", Espoo (Finland), 12-14 Sept, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI, Part 3/W52, pp. 20- 26

Barazzetti, L., Remondino, F., Scaioni, M., 2009. Combined use of photogrammetric and computer vision techniques for fully automated and accurate 3D modeling of terrestrial objects, Proc. SPIE 7447, Videometrics, Range Imaging, and Applications X, 74470M (3 September 2009). https://doi.org/10.1117/12.825638

Bentley Systems Incorporated, 2019. ContextCapture Software, Version 4.4.7.68. bentley.com (1 March 2021).

Bianchi, C.N., Pronzato, R., Cattaneo-Vietti, R., Benedetti-Cecchi, L., Morri, C., Pansini, M., Bavestrello, G., 2004. Hard bottoms, Biol. Mar. Medit., vol.11, 2004, pp.185-215

Boccardo P., Dequal S., Lingua A., Rinaudo F., 2001. True Digital Orthophoto for architectural and archaeological applications. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV-5/W1, 2001. International Workshop on Recreating the Past-Visualization and Animation of Cultural Heritage, 26 February – 1 March 2001, Ayutthaya, Thailand

Brown, M., Lowe, D.G., 2002. Invariant features from interest point groups. In British Machine Vision Conference, Cardiff, Wales, pp. 656-665.

Burns, J. H. R., Delparte, D., 2017. Comparison of commercial Structure-From-Motion Photogrammetry software used for underwater three-dimensional modelling of coral reef environments, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W3, 127-131, https://doi.org/10.5194/isprs-archives-XLII-2-W3-127-2017

Burns, J.H.R., Delparte, D., Gates, R.D., Takabayashi, M., 2015. Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3D ecological characteristics of coral reefs, PeerJ 3:e1077

Cannarozzo, R., Cucchiarini, L. Meschieri, W., 2012. Misure, rilievo, progetto - modulo T: Fotogrammetria, Zanichelli, Bologna, Italy.

Canny, J., 1986. A Computational Approach to Edge Detection, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 6, pp. 679-698, Nov. 1986, https://doi.org/10.1109/TPAMI.1986.4767851.

Capobianco, A., Gagliolo, S., Pallecchi, S., Sguerso, D., 2021. Analysis of historical evolution and present state of conservation of Regio VII, Insula 14 in Pompeii. Conference proceedings of ARQUEOLÓGICA 2.0 - 9th International Congress & 3rd GEORES - GEOmatics and pREServation

Carrasco, M., Mery, D., Concha, A., Velázquez, R., De Fazio, R., Visconti, P., 2021. An Efficient Point-Matching Method Based on Multiple Geometrical Hypotheses. Electronics, 10, 246. https://doi.org/10.3390/electronics10030246

Chen, X., Jiang, K., Zhu, Y., Wang, X., Yun, T., 2021. Individual Tree Crown Segmentation Directly from UAV-Borne LiDAR Data Using the PointNet of Deep Learning. Forests, 12(2), 131. doi.org/10.3390/f12020131

CloudCompare Development Team, 2021. CloudCompare Software, Version 2.12 alpha. cloudcompare.org (29 March 2021).

Cresson, R., 2020: Semantic segmentation of optical imagery. Deep Learning for Remote Sensing Images with Open Source Software, CRC Press

Di Pietra, V., Piras, M., Ruotsalainen, L., Rantanen, J., Malkamaki, T., 2018. A framework for multisensor positioning and navigation in underground mines. Poster presentation at 2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Nantes, France, http://ipin2018.ifsttar.fr/

Ding, Y., Zheng, X., Xiong, H., Zhang, Y., 2019. Semantic segmentation of indoor 3D point cloud with SLENet. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W13, 785–791. doi.org/10.5194/isprs-archives-XLII-2-W13-785-2019

Duda, R., Hart., P., 1973. Pattern classification and scene analysis, A Wiley-Interscience publication, pp. 271-272, https://doi.org/10.2307/2286028

Durrant-Whyte, H., Rye, D., Nebot, E., 1996. Localization of Autonomous Guided Vehicles. In: Giralt G., Hirzinger G. (eds) Robotics Research. Springer, London, pp. 613-625. https://doi.org/10.1007/978-1-4471-0765-1_69

Durrant-Whyte, H., Bailey, T., 2006. Simultaneous localization and mapping: part I, in IEEE Robotics & Automation Magazine, vol. 13, no. 2, pp. 99-110, https://doi.org/10.1109/MRA.2006.1638022

Eakin, C.M., Liu, G., Gomez, A.M., De La Cour, J.L., Heron, S.F., Skirving, W.J., Strong, A.E., 2017. Global coral bleaching 2014-2017, Reef Encounter, vol.31, pp.20-26

El-Ashmawy, K.L.A., 2018. Using direct linear transformation (DLT) method for aerial photogrammetry applications. Geodesy and Cartography, 44(3), 71-79. https://doi.org/10.3846/gac.2018.1629

Euler, L., 1775. Nova Methodus Motum Corporum Rigidorum Determinandi. Novi Commentari Academiae Scientiarum Imperialis Petropolitanae, Vol. 20, 1775, pp. 208–238.

Euler, L., 1770. Problema Algebraicum ob Affectiones Prorsus Singulares Memorabile. Novi Commentari Academiae Scientiarum Imperialis Petropolitanae, Vol. 15, 1770, pp. 75–106.

Fagandini, R., Federici, B., Ferrando, I., Gagliolo, S., Pagliari, D., Passoni D., Pinto, L., Rossi, L., Sguerso, D., 2017. Evaluation of the Laser Response of Leica Nova Multistation MS60 for 3D Modelling and Structural Monitoring. In International Conference on Computational Science and Its Applications, pp. 93-104. 10.1007/978-3-319-62401-3_8

Fangi, G., 1995. Note di fotogrammetria. Clua Edizioni Ancona

Fischler, M.A., Bolles, R.C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Association for Computing Machinery, vol. 24 nr. 6., pp. 381-395. 10.1145/358669.358692

Forlani, G. (1986). Sperimentazione del nuovo programma CALGE dell'ITM. Bollettino SIFET 1986, 2, 63–72.

Forsyth, D.A., Ponce, J., 2012. Computer Vision - A Modern Approach, Second Edition, Pitman.

Fraser, C.S., 1984. Network design considerations for non-topographic photogrammetry. Photogrammetric Engineering and Remote Sensing, vol. 50, no. 4, pp. 561-570, 1984.

Fraser, C.S., 1992. Photogrammetric measurement to one part in a million. Photogrammetric engineering and remote sensing, 58, 305-310.

Förstner, W., Wrobel, B.P., 2016. Photogrammetric Computer Vision: Statistics, Geometry, Orientation and Reconstruction. Geometry and Computing 11, Springer

Gabara, G., Sawicki, P., 2017. Study on 3D Point Clouds Accuracy of Elongated Object Reconstruction in Close Range – Comparison of Different Software, 10th International Conference Environmental Engineering, https://doi.org/10.3846/enviro.2017.188

Gagliolo, S., Sguerso, D., 2021. MAGO approach for semantic segmentation: the case study of UAVid benchmark dataset. In print on Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.

Gagliolo, S., Passoni, D., Federici, B., Ferrando, I., Sguerso, D., 2019. U.Ph.O and MAGO: two useful instruments in support of photogrammetric UAV survey, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W13, 289–296. 10.5194/isprs-archives-XLII-2-W13-289-2019

Gagliolo, S., Federici, B., Ferrando, I., Sguerso, D., 2019. MAGO: a new approach for orthophotos production based on adaptive mesh reconstruction, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W11, 533–538. 10.5194/isprs-archives-XLII-2-W11-533-2019

Gagliolo, S., 2019. Ortofoto ad alta risoluzione per individuazione di lesioni strutturali con “MAGO” (High resolution orthophotos for the recognition of structural lesions with “MAGO”), honoured with the "Premio Giovani Autori" (Young Authors Prize) during the 64° Convegno SIFET. Bollettino SIFET 2019, vol. 1, abstract in English

Gagliolo, S., Fagandini, R., Passoni, D., Federici, B., Ferrando, I., Pagliari, D., Pinto, L., Sguerso, D., 2018. Parameter optimization for creating reliable photogrammetric models in emergency scenarios, Applied Geomatics, 10 (4), pp. 501-514. 10.1007/s12518-018-0224-4

Gagliolo, S., Ausonio, E., Federici, B., Ferrando, I., Passoni, D., Sguerso, D., 2018. 3D cultural heritage documentation: A comparison between different photogrammetric software and their products, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2, 347–354. 10.5194/isprs-archives-XLII-2-347-2018

Gagliolo, S., Fagandini, R., Federici, B., Ferrando, I., Passoni, D., Pagliari, D., Pinto, L., Sguerso, D., 2017. Use of UAS for the conservation of historical buildings in case of emergencies. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XLII-5/W1, 81-88. 10.5194/isprs-Archives-XLII-5-W1-81-2017

Gini, R., Pagliari, D., Passoni, D., Pinto, L., Sona, G., and Dosso, P., 2013. UAV Photogrammetry: block triangulation comparisons. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XL-1/W2, 157-162, https://doi.org/10.5194/isprsarchives-XL-1-W2-157-2013

Gomarasca, M.A., 2009. Basics of Geomatics. Springer Netherlands. https://doi.org/10.1007/978-1-4020-9014-1

Ghosh, S. K., 2005. Fundamentals of computational photogrammetry. New Delhi: Concept Publishing Co.

Grafarend, E.W., 1974. Optimization of geodetic networks. Bolletino di Geodesia a Science Affini, 33(4), 351-406.

GRASS Development Team, 2018. Geographic Resources Analysis Support System (GRASS) Software, Version 7.4. Open Source Geospatial Foundation. grass.osgeo.org (1 May 2021).

Grün, A., 1978. Progress in photogrammetric point determination by compensation of systematic errors and detection of gross errors. Inform. Relative to Cartography and Geodesy. Ser. 2: Transl. p 113-140(SEE N 80-12455 03-43).

Grün, A., Wang, X., 1998. CC-Modeler: A Topology Generator for 3-D City Models. ISPRS Journal of Photogrammetry and Remote Sensing 53 (5): 286–295.

Grün, A., 1980. Precision and reliability aspects in Close-Range photogrammetry. Invited paper, XIV Congress of ISP, Comm. V., Hamburg.

Grün, A., M. Behnisch, Kohler, N., 2009. Perspectives in the Reality-Based Generation, n D Modelling, and Operation of Buildings and Building Stocks. Building Research & Information 37 (5–6): 503–519.

Haala, N., Cavegn, S., 2016. High density aerial image matching: state-of-the-art and future prospects, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLI-B4, 625-630, https://doi.org/10.5194/isprs-archives-XLI-B4-625-2016

Haala, N., 2013. The Landscape of Dense Image Matching Algorithms. In Photogrammetric Week ’ 13; Fritsch, D., Ed.; Wichmann: Stuttgart, Germany, 2013; pp. 271–284.

Habib A. F., Kim E., Kim C., 2007. New Methodologies for True Orthophoto Generation. Photogrammetric Engineering & Remote Sensing, Vol. 73, No. 1, January 2007, pp. 025–036.

Hamilton, W. R., 1844. On Quaternions; or a New System of Imaginaries in Algebra. Philosophical Magazine, Vol. 25, 1844, pp. 489–495.

Harris, C., Stephens, M., 1988. A combined corner and edge detector, in Proc. of Fourth Alvey Vision Conference, 1988, pp.147-151

Heron, S.F., Eakin, C.M., Douvere, F., 2017. Impacts of climate change on World Heritage Coral Reefs: A first global scientific assessment, UNESCO World Heritage Centre, Paris, France, 2017

Honkavaara, E., Markelin, L., Rosnell, T., Nurminen, K., 2012. Influence of solar elevation in radiometric and geometric performance of multispectral photogrammetry, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 67, pp. 13-26, https://doi.org/10.1016/j.isprsjprs.2011.10.001

Hough, P.V.C., 1962. A Method and Means for Recognizing Complex Patterns, U.S. Patent 3,069,654, December 18, 1962.

Huang, S., Nex, F., Lin, Y., Yang, M. Y., 2019. Semantic segmentation of building in airborne images. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W13, 35–42. doi.org/10.5194/isprs-archives-XLII-2-W13-35-2019

Hughes, T.P., Anderson, K.D., Connolly, S.R., Heron, S.F., Kerry, J.T., Lough, J.M., Wilson, S.K., 2018. Spatial and temporal patterns of mass bleaching of corals in the Anthropocene, Science, vol.359, pp.80-83

Hughes, T.P., Kerry, J.T., Álvarez-Noriega, M., Álvarez-Romero, J.G., Anderson, K.D., Baird, A.H., Wilson, S.K., 2017. Global warming and recurrent mass bleaching of corals, Nature, vol.543, pp.373-377

IGN (French National Geographic Institute) and ENSG (French national school for geographic sciences), 2019. MicMac Software, Version 1.0.beta11. micmac.ensg.eu (1 March 2021)

Iturralde, D., Azurdia-Meza, C., Krommenacker, N., Soto, I., Ghassemlooy, Z., Becerra, N., 2014. A new location system for an underground mining environment using visible light communications. IEEE, 9th International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP 2014), pp. 1165-1169

Jin, Y.-H., Ko, K.-W., Lee, W.-H., 2018. An Indoor Location-Based Positioning System Using Stereo Vision with the Drone Camera. Mobile Information Systems, vol. 2018, Article ID 5160543, 13 pages, 2018. 10.1155/2018/5160543

Kaasalainen, S., Malkamaki, T., Ilinca, J., Ruotsalainen, L., 2018. Multispectral Terrestrial Laser Scanning: new developments and applications. IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 8737-8740

Kahl, F., Agarwal, S., Chandraker, M.K. Kriegman, D., Belongie, S., 2008. Practical Global Optimization for Multiview Geometry. International Journal of Computer Vision 79, 271–284. https://doi.org/10.1007/s11263-007-0117-1

Kanatani, K., 1990. Group Theoretical Methods in Image Understanding. New York: Springer.

Kempf, C., Tian, J., Kurz, F., d'Angelo, P., Reinartz, P., 2019. Local versus global variational approaches to enhance watershed transformation based individual tree crown segmentation of digital surface models from 3k optical imagery. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W13, 43–50. doi.org/10.5194/isprs-archives-XLII-2-W13-43-2019

Krauss, K., 1993. Photogrammetry Vol. I., Fundamentals and Standard Processes, Der. Dümmler Verlag, Bonn, Germany, ISBN 3-427-78684-6

Lafferty, J.D., McCallum, A., Pereira, F.C.N., 2001: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), Williamstown, MA, USA, 28 June–1 July 2001; pp. 282–289.

Lague, D., Brodu, N., Leroux, J., 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: application to the Rangitikei canyon (N-Z). ISPRS Journal of Photogrammetry and Remote Sensing 82, pp. 10-26

Lasagna, R., Gattorna, I., Albertelli, G., Morri, C., Bianchi, C.N., 2006. Substrate heterogeneity and relation with coral recruitment in coral reefs of the Maldives (Indian Ocean), Biol. Mar. Medit., vol.13, pp.88-89

Leavers, V. F., 1992. Preprocessing, Shape Detection in Computer Vision Using Hough Transform, Springer, pp. 39-64. 10.1007/978–1–4471–1940–1, ISBN 978-3-540-19723-2

Lesak, M., 2019. Odometry for autonomous navigation in GPS denied environments. 2019 - Mines Theses & Dissertations, Colorado School of Mines, https://hdl.handle.net/11124/173044

Li, B., Zhao, K., Saydam, S., Rizos, C., Wang, Q., Wang, J., 2018. Positioning technologies for underground mines. Far East Journal of Electronics and Communications; v. 18; no. 6; pp. 871 - 893; ISSN 0973-7006. http://handle.unsw.edu.au/1959.4/unsworks_55776

Li, T., Jiang, C., Bian, Z., Wang, M., Niu, X., 2020. A Review of True Orthophoto Rectification Algorithms, IOP Conference Series: Materials Science and Engineering, IOP Publishing, vol. 780, https://doi.org/10.1088/1757-899x/780/2/022035

Lillesand, T., R. W. Kiefer, Chipman, J., 2014. Remote Sensing and Image Interpretation. New York: John Wiley.

Liu, Z., Liu, Y., Chen, B., Yang, C., Chen, Y. P., Chen, R., 2018. Smooth point matching based on SIFT, 2018 International Conference on Cloud Computing, Big Data and Blockchain (ICCBB), pp. 1-4, https://doi.org/10.1109/ICCBB.2018.8756361

Lowe, D.G., 1999. Object recognition from local scale-invariant features. In International Conference on Computer Vision, Corfu, Greece, pp. 1150-1157.

Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60, 91–110 (2004). 10.1023/B:VISI.0000029664.99615.94

Lyu, Y., Vosselman, G., Xia, G.S., Yilmaz, A., Yang, M.Y., 2020: UAVid: A semantic segmentation dataset for UAV imagery, ISPRS Journal of Photogrammetry and Remote Sensing, 165, 108 – 119. doi.org/10.1016/j.isprsjprs.2020.05.009

Lu, D., P. Mausel, E. Brondizio, Moran, E., 2004. Change Detection Techniques. International Journal of Remote Sensing 25 (12): 2365–2401.

Lösch, R., Grehl, S., Donner, M., Buhl, C., Jung, B., 2018. Design of an Autonomous Robot for Mapping, Navigation, and Manipulation in Underground Mines. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1407-1412, 10.1109/IROS.2018.8594190

Mach, L., 2009. Semi-automatic system for reconstruction of 3D scenes. Bachelor Thesis at Charles University in Prague, Faculty of Mathematics and Physics

Malkamäki, T., Kaasalainen, S., Ilinca, J., 2018. Improving the Full Waveform Digitizing in an Eight Channel Portable Hyperspectral Lidar. Light, Energy and the Environment 2018 (E2, FTS, HISE, SOLAR, SSL), OSA Technical Digest (Optical Society of America, 2018), paper HM4C.4

Malkamäki, T., Kaasalainen, S., Ilinca, J., 2019. Portable hyperspectral lidar utilizing 5 GHz multichannel full waveform digitization. Opt. Express 27, A468-A480

Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M., Stilla, U., 2018. Classification with an edge: improving semantic image segmentation with boundary detection. ISPRS Journal of Photogrammetry and Remote Sensing 135, 158-172. doi.org/10.1016/j.isprsjprs.2017.11.009

Marr, D., Hildreth, E., 1980. Theory of edge detection, Proceedings of the Royal Society B.207187–217, http://doi.org/10.1098/rspb.1980.0020

Martinez-Soltero, G., Alanis, A.Y., Arana-Daniel, N., Lopez-Franco, C., 2020: Semantic Segmentation for Aerial Mapping. *Mathematics*, 8, 1456. doi.org/10.3390/math8091456

Montefalcone, M., Morri, C., Bianchi, C.N., 2018. Long term change in bioconstruction potential of Maldivian coral reefs following extreme climate anomalies, Glob. Change Biol., vol.24, pp.5629-5641

Morri, C., Montefalcone, M., Lasagna, R., Gatti, G., Rovere, A., Parravicini, V., Baldelli, G., Colantoni, P., Bianchi, C.N., 2015. Through bleaching and tsunami: coral reef recovery in the Maldives, Mar. Pollut. Bull., vol.98, 2015, pp.188-200

Ngadiman, N., Kaamin, M., Sahat, S., Mokhtar, M., Ahmad, N.F.A., Kadir, A.A., Razali, S.N.M., 2018. Production of orthophoto map using UAV photogrammetry: A case study in UTHM Pagoh campus, in Proceedings of the 3rd International Conference on Applied Science and Technology (ICAST'18), 2018, vol. 2016, no. 1. doi:10.1063/1.5055514

Niederheiser, R., Mokroš, M., Lange, J., Petschko, H., Prasicek, G., and Elberink, S. O., 2016. Deriving 3D point clouds from terrestrial photographs – comparison of different sensors and software, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLI-B5, 685-692, https://doi.org/10.5194/isprs-archives-XLI-B5-685-2016

Niemeier, W., 1982. Design, diagnosis and optimization of monitoring networks in engineering surveying. In Invited Paper, Centennial Convention of CIS, Ottawa.

Nikhil, R.P., Sankar, K.P., 1993. A review on image segmentation techniques, Pattern Recognition, Vol. 26, Issue 9, 1993, pp. 1277-1294, https://doi.org/10.1016/0031-3203(93)90135-J

Nocerino, E., Neyer, F., Gruen, A., Troyer, M., Menna, F., Brooks, A., Capra, A., Castagnetti, C., Rossi, P., 2019. Comparison of diver-operated underwater photogrammetric systems for coral reef monitoring, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W10, 2019, pp.143-150

Ondrašovič, M., Tarábek, P., 2021. Homography Ranking Based on Multiple Groups of Point Correspondences, Sensors, vol. 21, nr. 17, https://doi.org/10.3390/s21175752

OpenCV Development Team, 2019. OpenCV 4.1.2 Source Library. opencv.org (1 March 2021)

OpenStreetMap® Foundation (OSMF), 2021. OpenStreetMap® QuickMapServices available in QGIS 3. openstreetmap.org

Ouyang, S., Li, Y., 2021: Combining Deep Semantic Segmentation Network and Graph Convolutional Neural Network for Semantic Segmentation of Remote Sensing Imagery. *Remote Sens.*, 13, 119. doi.org/10.3390/rs13010119

Pan, X., Zhao, J., Xu, J., 2020: An End-to-End and Localized Post-Processing Method for Correcting High-Resolution Remote Sensing Classification Result Images. Remote Sens., 12, 852. doi.org/10.3390/rs12050852

Pagliari, D., Rossi, L., Passoni, D., Pinto, L., De Michele, C., Avanzi, F., 2017. Measuring the volume of flushed sediments in a reservoir using multi-temporal images acquired with UAS. Geomatics, Natural Hazards and Risk, 8(1), 150-166.

Passoni, D., 2019. Innovative Tools For Planning, Analysis, and Management of UAV Photogrammetric Surveys, PhD thesis.

Passoni, D., Federici, B., Ferrando, I., Gagliolo, S., Sguerso, D., 2018. The estimation of precisions in the planning of UAS photogrammetric surveys, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2, pp. 837-843. 10.5194/isprs-archives-XLII-2-837-2018

Pavelka, K., Řezníček, J., Bílá, Z., Prunarová, L., 2013. Non Expensive 3D Documentation and Modelling of Historical Object and Archaeological Artefacts by Using Close Range Photogrammetry, Geoinformatics FCE CTU 10. 10.14311/gi.10.5

Perfetti, L., Polari, C., Fassi, F., Troisi, S., Baiocchi, V., Del Pizzo, S., Giannone, F., Barazzetti, L., Previtali, M., Roncoroni, F., 2018. Fisheye Photogrammetry to Survey Narrow Spaces in Architecture and a Hypogea Environment, in Latest Developments in Reality-Based 3D Surveying and Modelling, Remondino, F., Georgopoulos, A., González-Aguilera, D., Agrafiotis, P., Eds.; MDPI: Basel, Switzerland, 2018; pp. 3–28. 10.3390/books978-3-03842-685-1-1

Perry, C.T., Morgan, K.M., 2017. Bleaching drives collapse in reef carbonate budgets and reef growth potential on southern Maldives reefs, Sci. Reports, vol.7, 2017, 40581

Pix4D SA, 2019. Pix4Dmapper© Software, Version 4.1.25. pix4d.com (1 March 2021).

QGIS Development Team, 2019. QGIS Software, Version 3.4. Open Source Geospatial Foundation. https://www.osgeo.org/projects/qgis/ (1 May 2019).

QGIS Development Team, 2020. QGIS Software, Version 3.12. Open Source Geospatial Foundation. https://www.osgeo.org/projects/qgis/ (1 May 2021).

Qin, R., 2015. 3D Change Detection in an Urban Environment with Multi-Temporal Data. Ph.D Thesis, Swiss Federal Institute of Technology, ETH, Zürich.

Qin, R., J. Tian, Reinartz P., 2016. 3D Change Detection – Approaches and Applications. ISPRS Journal of Photogrammetry and Remote Sensing 122: 41–56.

Qin, R., and Grün, A., 2020. The Role of Machine Intelligence in Photogrammetric 3D Modeling – An Overview and Perspectives. International Journal of Digital Earth 1–17. https://doi.org/10.1080/17538947.2020.1805037

Qiu, M., Zhang, Y., 2015. Feature guided multi-window area-based matching method for urban remote sensing stereo pairs, 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4510-4513, https://doi.org/10.1109/IGARSS.2015.7326830

Re, C., Roncella, R., Forlani, G., Cremonese, G., Naletto, G., 2014. Evaluation of an Area-Based matching algorithm with advanced shape models, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XL-4, 215–221, https://doi.org/10.5194/isprsarchives-XL-4-215-2014

Remondino, F., Spera, M. G., Nocerino, E., Menna, F., Nex, F., 2014. State of the art in high density image matching, The Photogrammetric record, Volume 29, Issue 146, June 2014, pp. 144–166, DOI: 10.1111/phor.12063

Ren, Z., Wang, L., Bi, L., 2019. Robust GICP-Based 3D LiDAR SLAM for Underground Mining Environment. Sensors 2019, 19, 2915

Rodriguez, O., 1840. Des lois géometriques qui regissent les deplacements d'un système solide independament des causes qui peuvent les produire. Journal de mathématiques pures et appliquées 1 (5), 380–440.

Rossi, P., Castagnetti, C., Capra, A., Brooks, A.J., Mancini, F., 2019. Detecting change in coral reef 3D structure using underwater photogrammetry: critical issues and performance metrics, Applied Geomatics, 2019, pp. 1-15

Sai, S.S., Tjahjadi, M.E., Rokhmana, C.A., 2019. Geometric Accuracy Assessments of Orthophoto Production from UAV Aerial Images. KnE Engineering, 4(3), 333–344. https://doi.org/10.18502/keg.v4i3.5876

Schmitz, M., Huang, H., Mayer, H., 2019. Comparison of training strategies for convnets on multiple similar datasets for facade segmentation. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W13, 111–117. doi.org/10.5194/isprs-archives-XLII-2-W13-111-2019

Schöning J., Heidemann G., 2015. Evaluation of Multi-view 3D Reconstruction Software. In: Azzopardi G., Petkov N. (eds) Computer Analysis of Images and Patterns. CAIP 2015. Lecture Notes in Computer Science, vol 9257. Springer, Cham. 10.1007/978-3-319-23117-4_39

Sguerso, D., Azzola, A., Bianchi, C.N., Federici, B., Gagliolo, S., Morri, C., Montefalcone, M., 2020. 3D photogrammetric surveys on coral reefs in the Maldives, 2019 IMEKO TC19 International Workshop on Metrology for the Sea: Learning to Measure Sea Health Parameters, MetroSea 2019, Genoa, Italy, 3-5 October 2019, pp. 292-297

Sguerso, D., Bianchi, C.N., Federici, B., Ferrando, I., Gagliolo, S., Montefalcone, M., Morri, C., Passoni, D., 2019. Rilevamento fotogrammetrico 3D delle scogliere coralline delle Maldive, poster presentation at 64° Convegno Sifet, Venezia 19-21 June 2019, extended abstract in Italian in conference proceedings. Honoured with the Poster Award: Marine Surveying

Shan, J., Z. Hu, P. Tao, L. Wang, S. Zhang, Ji, S., 2020. Toward a Unified Theoretical Framework for Photogrammetry. Geo-spatial Information Science 23 (1): 75–86.

Shao, F., Liu, Z., Jubai, A., 2021. Feature Point Matching Based on Local Relative Velocity Consensus. Journal of Physics: Conference Series, IOP Publishing, vol. 1871, https://doi.org/10.1088/1742-6596/1871/1/012055

Shao, W., Vijayarangan, S., Li, C., Kantor, G., 2019. Stereo Visual Inertial LiDAR Simultaneous Localization and Mapping. CoRR, abs/1902.10741, 2019, http://arxiv.org/abs/1902.10741

Sheta, B., Elhabiby, M., El-Sheimy, N., 2012. Assessments of Nonlinear Least Squares methods for UAV vision based navigation. ASPRS 2012 Annual Conference Sacramento, California, March 19-23, 2012. http://www.asprs.org/a/publications/proceedings/Sacramento2012/files/Sheta.pdf

Soille, P., 2004. Erosion and Dilation. In: Morphological Image Analysis. Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-662-05088-0_3

Sothe, C., De Almeida, C. M., Schimalski, M. B., La Rosa, L. E. C., Castro, J. D. B., Feitosa, R. Q., Dalponte, M., Lima, C. L., Liesenberg, V., Miyoshi, G. T., Tommaselli, A. M. G., 2020. Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data. GIScience & Remote Sensing, 57(3), 369-394. doi.org/10.1080/15481603.2020.1712102

Stachniss, C., Leonard, J.J., Thrun, S., 2016. Simultaneous Localization and Mapping. In: Siciliano B., Khatib O. (eds) Springer Handbook of Robotics. Springer Handbooks. Springer, Cham. https://doi.org/10.1007/978-3-319-32552-1_46

Svensk, J., 2017. Evaluation of aerial image stereo matching methods for forest variable estimation, Master of Science Thesis in Computer Vision Department of Electrical Engineering, Linköping University

Thrybom, L., Neander, J., Hansen, E., Landernas, K., 2015. Future Challenges of Positioning in Underground Mines. IFAC-PapersOnLine, Volume 48, Issue 10, 2015, Pages 222-226, ISSN 2405-8963, https://doi.org/10.1016/j.ifacol.2015.08.135

Tommaselli, A. M. G., Campos, M. B., Castanheiro, L. F., Honkavaara, E., 2019. A feasibility study on incremental bundle adjustment with fisheye images and low-cost sensors, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W18, 167–171. 10.5194/isprs-archives-XLII-2-W18-167-2019

Triggs B., McLauchlan P.F., Hartley R.I., Fitzgibbon A.W., 2000. Bundle Adjustment — A Modern Synthesis. In: Triggs B., Zisserman A., Szeliski R. (eds) Vision Algorithms: Theory and Practice. IWVA 1999. Lecture Notes in Computer Science, vol 1883. Springer, Berlin, Heidelberg

Torlegård, K., 1980. On accuracy improvement in Close-Range photogrammetry, Proceedings of the Industrial and Engineering Survey Conference, London.

Ullman, S., 1979. The Interpretation of Structure from Motion. Proceedings of the Royal Society of London Series B, 203(1153), 405-426. doi.org/10.1098/rspb.1979.0006

Wang Q., Yan L., Sun Y., Cui X., Mortimer H., Li Y., 2018. True orthophoto generation using line segment matches. The Photogrammetric Record, 33(161), pp. 113-130. DOI: 10.1111/phor.12229. 2018

Wojtas, A. M., 2010. Off-the-shelf close-range photogrammetric software for cultural heritage documentation at Stonehenge, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVIII, Part 5. Commission V Symposium, Newcastle upon Tyne, UK

Wu, C., 2013. Towards Linear-time Incremental Structure From Motion, 2013 International Conference on 3D Vision - 3DV 2013, pp. 127-134, doi: 10.1109/3DV.2013.25

Wu, C., Agarwal, S., Curless, B., Seitz, S.M., 2011. Multicore Bundle Adjustment, CVPR 2011, pp. 3057-3064, doi: 10.1109/CVPR.2011.5995552.

Zhang, J., Singh, S., 2018. Laser-visual-inertial odometry and mapping with high robustness and low drift. Journal of Field Robotics, August 2018

Zhengyu, X., Joohee, K., 2020: Mixed spatial pyramid pooling for semantic segmentation. *Applied Soft Computing*, 91, 106209. doi.org/10.1016/j.asoc.2020.106209

Zhou, K., Chen, Y., Smal, I., Lindenbergh, R., 2019. Building segmentation from airborne VHR images using mask R-CNN. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W13, 155–161. doi.org/10.5194/isprs-archives-XLII-2-W13-155-2019

Zoltan, K., Josiane, Z., 2012: Markov Random Fields in Image Segmentation. Collection Foundation and Trends in Signal Processing. Now Editor, World Scientific, pp.164.

ACKNOWLEDGEMENTS

I can safely say that these three (plus one) years have been the hardest yet the most wonderful period in my life, up to now. The experience of PhD helped me to grow wiser, to know myself (γνῶθι σεαυτόν), to find my way, to travel around the world, and to be brave in facing life. This wouldn't be the same without my extraordinary supervisor, prof. Domenico Sguerso, an enthusiast leader who is able to transfer his instinctive love for the job and the subject.

Thanks to prof. Bianca Federici and Dr. Ilaria Ferrando to be an example to follow and a support to learn.

Thanks to Daniele Passoni, Stefania Viaggio, Elena Ausonio, Roberta Fagandini, Lorenzo Benvenuto, who shared with me at least a short interval during these years.

Thanks to Tiziano Cosso, who invested in me when I was just a rookie, and to Roberto Marzocchi and Maria Federici from Gter srl Innovazione in Geomatica Gnss e Gis.

Thanks to the staff of DICCA Department and to all the colleagues in University of Genoa with whom I had the opportunity to interact.

I have been very grateful to know competent colleagues also all around Italy and Europe; in particular I wish to thank the team from Geodesy and Geomatics Area at Polytechnic of Milan, who proficiently contributed in my education and with whom I had the pleasure to cooperate as teaching assistant and for paper writing, and the team from the Finnish Geospatial Research Institute, in the person of the head of the NAVI department prof. Sanna Kaasalainen, to host me for four months in their amazing country and company.

A special remark on my Finnish colleagues Tuomo Malkamäki and Toni Hammarberg for their irreplaceable contribution in my expertise growth, fundamental for my work progression. It wouldn't be the same without the enthusiasm of Vincenzo Di Pietra, who shared with me his knowledge and love coming from his opportunity to visit FGI and Finland, convincing me to undertake this wonderful experience in turn.

An enormous thanks to my husband Daniele, who always believed in me and who ignited the spark to take up this career. I hope our sacrifices won't be vane in the future.

The last but not least thought is for my family, which I always hope to make proud of me, and my friends, who are close to me with their love and esteem.

APPENDIX – COMPLETE COLLECTION OF MAPS RELATED TO U.PH.O. TEST CASE NR.3

The appendix is dedicated to the report of images about the U.Ph.O. case study nr.3. In particular, the maps related to the third test are shown exhaustively, together with related tables. Comments on this topic are available at paragraph 4.2.2.

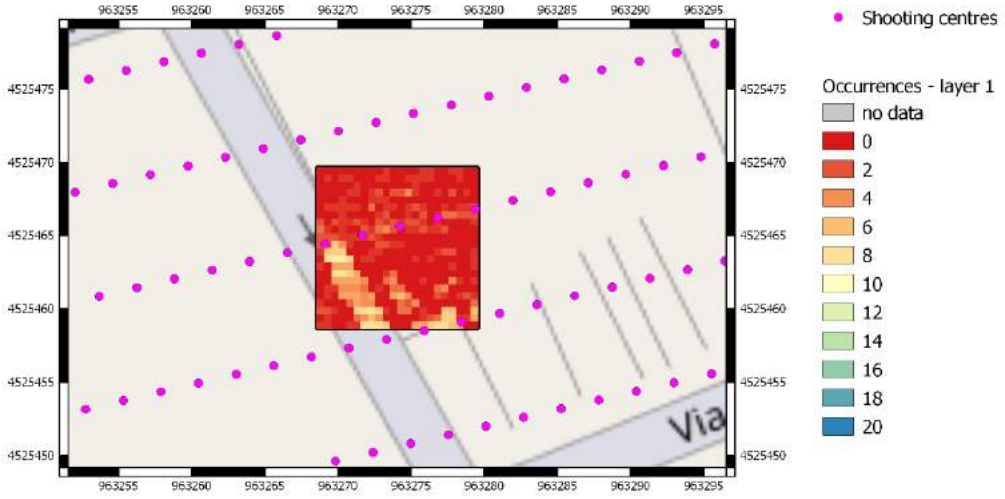


Figure 1 Occurrence map for the case study nr.3, layer 1.

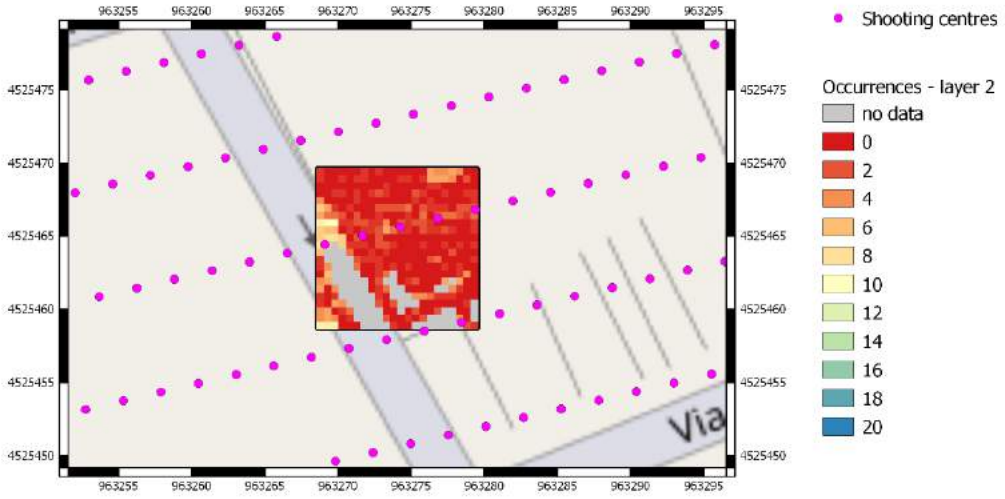


Figure 2 Occurrence map for the case study nr.3, layer 2.

Table 1 Classification of occurrence map: cell count and percentage cover for the observed classes in case study nr.3, layers 1 and 2.

Class	Nr. of occurrences	Layer 1		Layer 2	
		Cell count	% cover	Cell count	% cover
no data		0	0.0%	76	no data
0	0	261	53.9%	213	52.2%
1	1-2	140	28.9%	128	31.4%
2	3-4	35	7.2%	33	8.1%
3	5-6	23	4.8%	19	4.7%
4	7-8	23	4.8%	9	2.2%
5	9-10	2	0.4%	6	1.5%

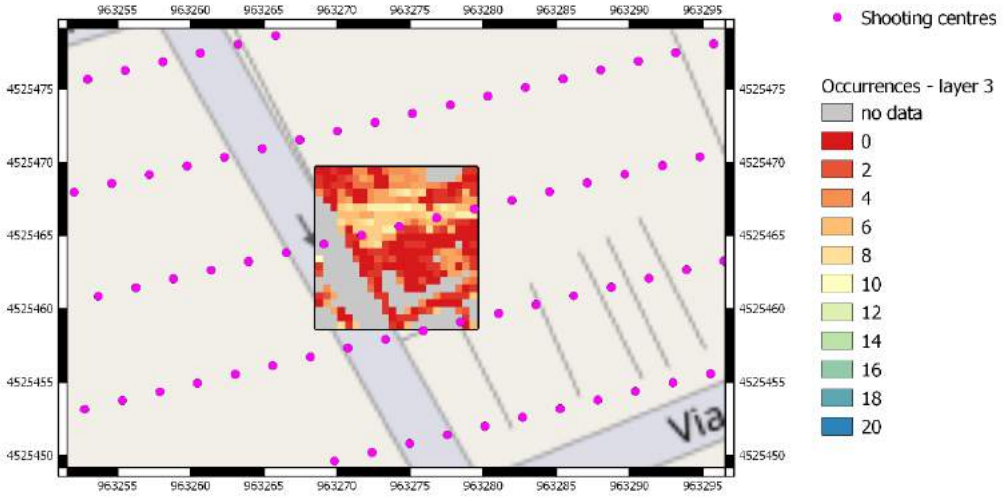


Figure 3 Occurrence map for the case study nr.3, layer 3.

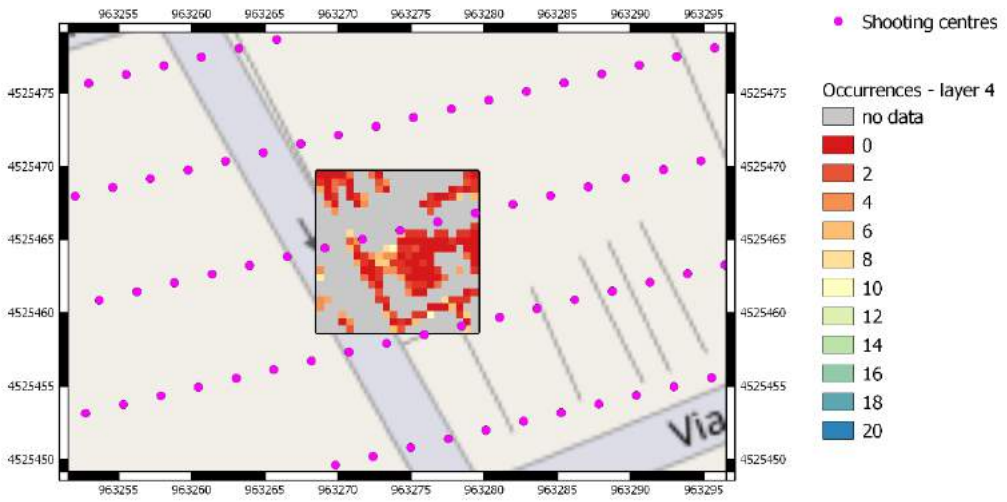


Figure 4 Occurrence map for the case study nr.3, layer 4.

Table 2 Classification of occurrence map: cell count and percentage cover for the observed classes in case study nr.3, layers 3 and 4.

Class	Nr. of occurrences	Layer 3		Layer 4	
		Cell count	% cover	Cell count	% cover
no data		142	29.3%	301	62.2%
0	0	108	31.6%	74	40.4%
1	1-2	81	23.7%	59	32.2%
2	3-4	43	12.6%	24	13.1%
3	5-6	56	16.4%	14	7.7%
4	7-8	39	11.4%	9	4.9%
5	9-10	15	4.4%	3	1.6%

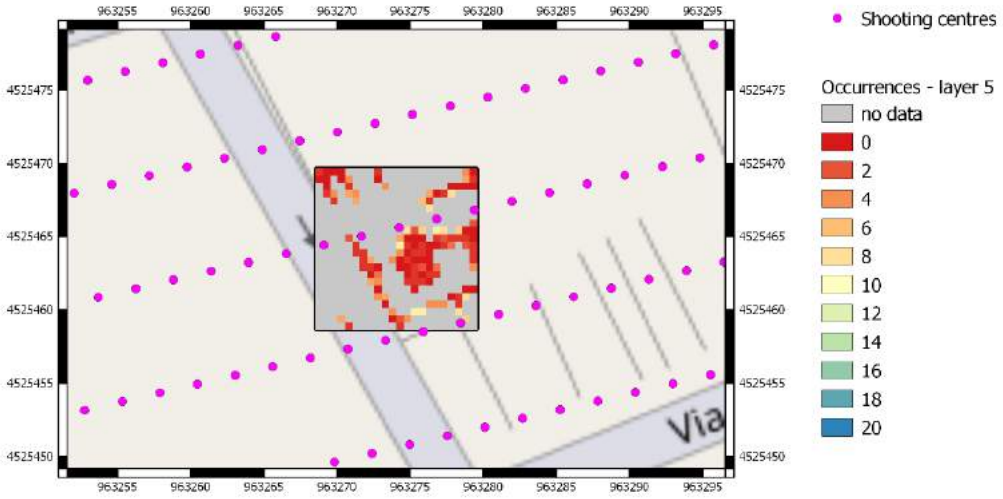


Figure 5 Occurrence map for the case study nr.3, layer 5.

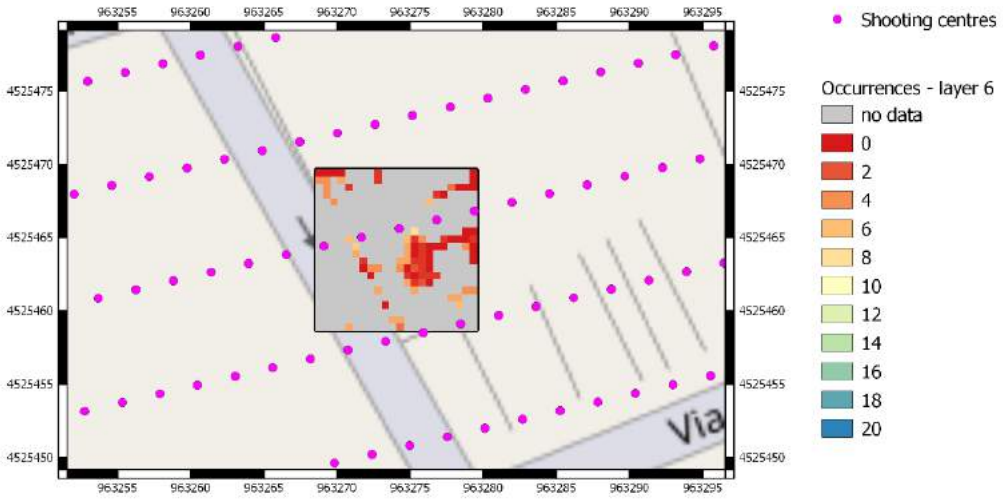


Figure 6 Occurrence map for the case study nr.3, layer 6.

Table 3 Classification of occurrence map: cell count and percentage cover for the observed classes in case study nr.3, layers 5 and 6.

Class	Nr. of occurrences	Layer 5		Layer 6	
		Cell count	% cover	Cell count	% cover
no data		354	73.1%	398	82.2%
0	0	44	33.8%	32	37.2%
1	1-2	45	34.6%	20	23.3%
2	3-4	18	13.8%	14	16.3%
3	5-6	10	7.7%	18	20.9%
4	7-8	9	6.9%	2	2.3%
5	9-10	4	3.1%	0	0.0%

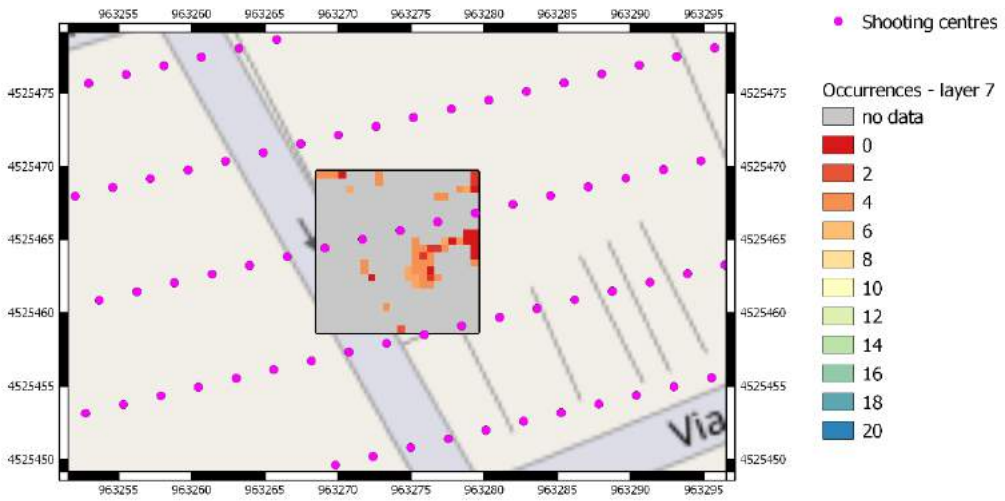


Figure 7 Occurrence map for the case study nr.3, layer 7.

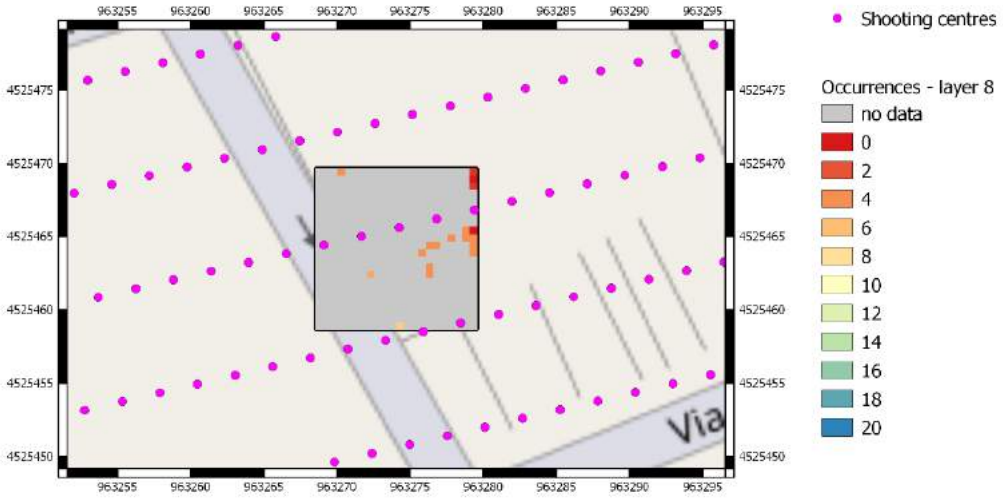


Figure 8 Occurrence map for the case study nr.3, layer 8.

Table 4 Classification of occurrence map: cell count and percentage cover for the observed classes in case study nr.3, layers 7 and 8.

Class	Nr. of occurrences	Layer 7		Layer 8	
		Cell count	% cover	Cell count	% cover
no data		431	89.1%	466	96.3%
0	0	11	20.8%	2	11.1%
1	1-2	7	13.2%	2	11.1%
2	3-4	28	52.8%	12	66.7%
3	5-6	7	13.2%	1	5.6%
4	7-8	0	0.0%	1	5.6%
5	9-10	0	0.0%	0	0.0%

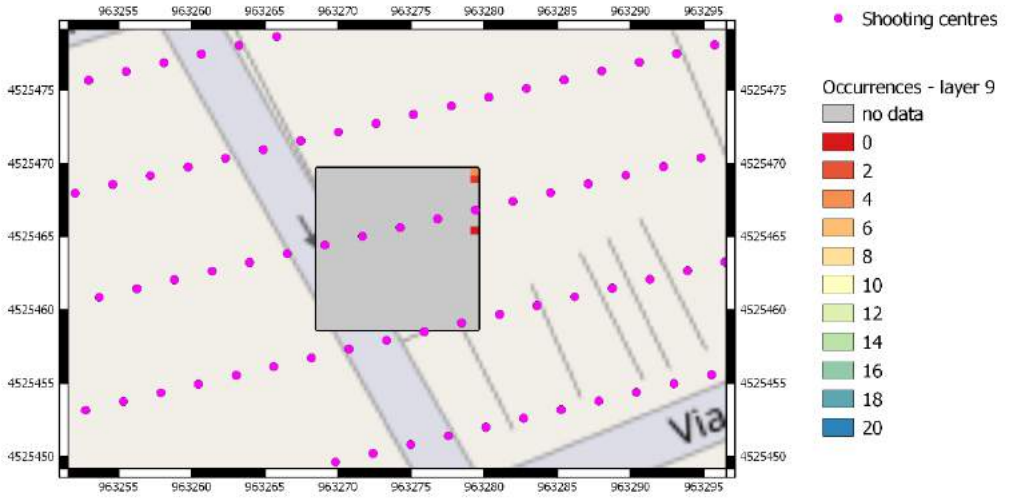


Figure 9 Occurrence map for the case study nr.3, layer 9.

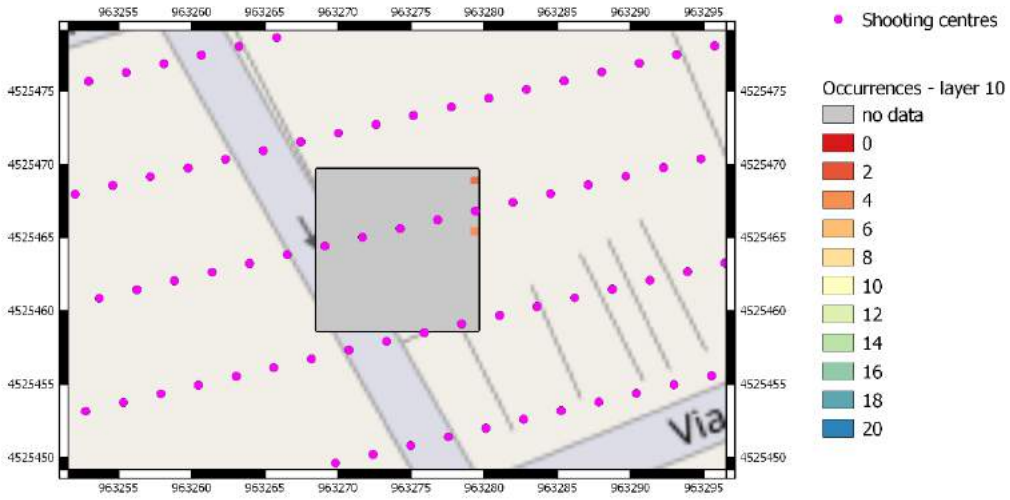


Figure 10 Occurrence map for the case study nr.3, layer 10.

Table 5 Classification of occurrence map: cell count and percentage cover for the observed classes in case study nr.3, layers 9 and 10.

Class	Nr. of occurrences	Layer 9		Layer 10	
		Cell count	% cover	Cell count	% cover
no data		481	99.4%	482	99.6%
0	0	1	33.3%	0	0.0%
1	1-2	1	33.3%	0	0.0%
2	3-4	1	33.3%	2	100.0%
3	5-6	0	0.0%	0	0.0%
4	7-8	0	0.0%	0	0.0%
5	9-10	0	0.0%	0	0.0%

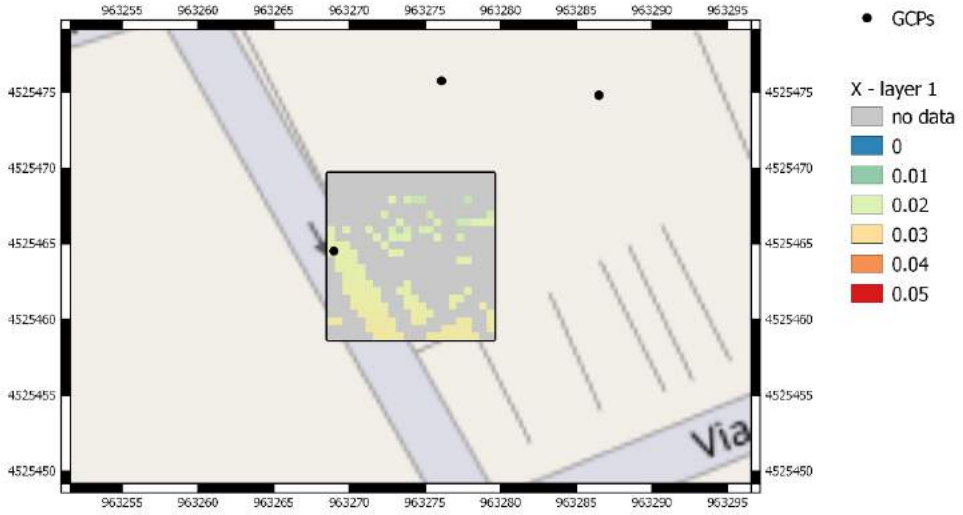


Figure 11 Accuracy [m] map for the X component of case study nr.3, layer 1.
Black dots represent GCPs, acquired by means of GPS in NRTK.

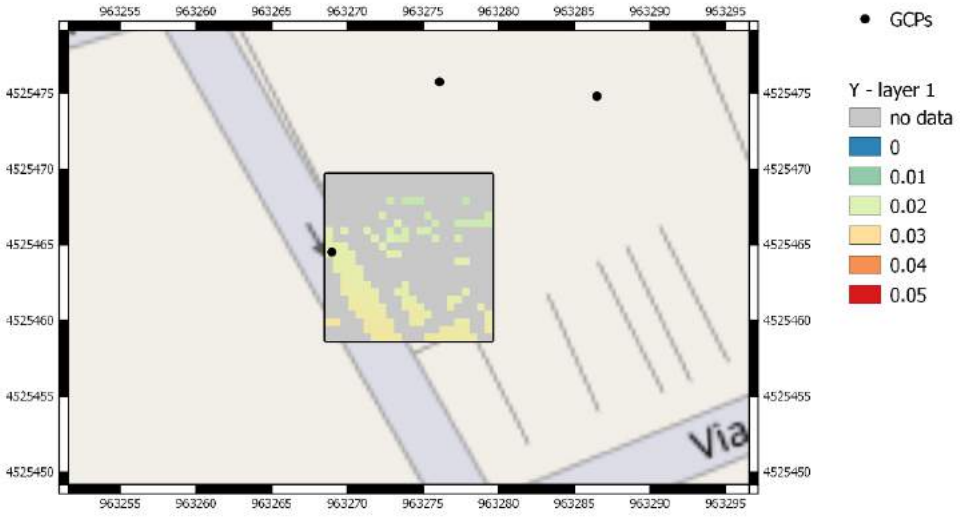


Figure 12 Accuracy [m] map for the Y component of case study nr.3, layer 1. Black dots represent GCPs, acquired by means of GPS in NRTK.

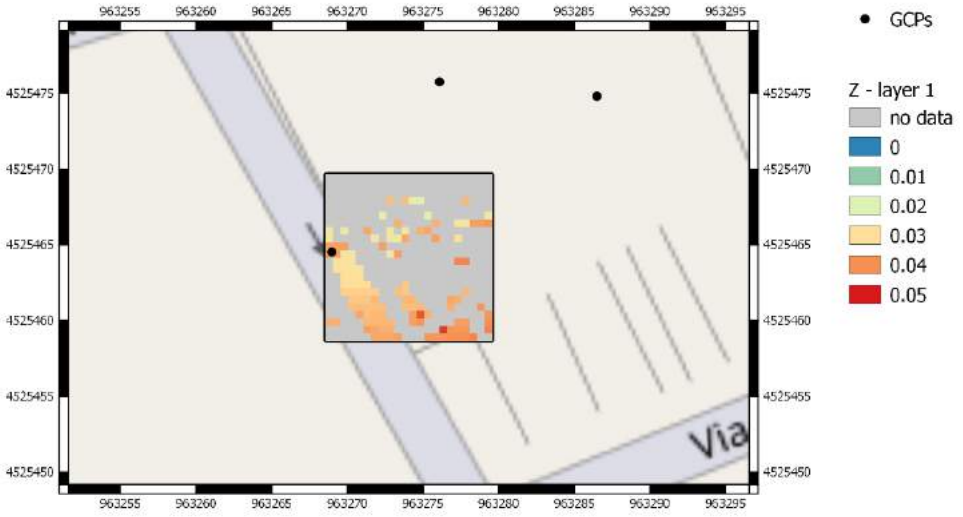


Figure 13 Accuracy [m] map for the Z component of case study nr.3, layer 1. Black dots represent GCPs, acquired by means of GPS in NRTK.

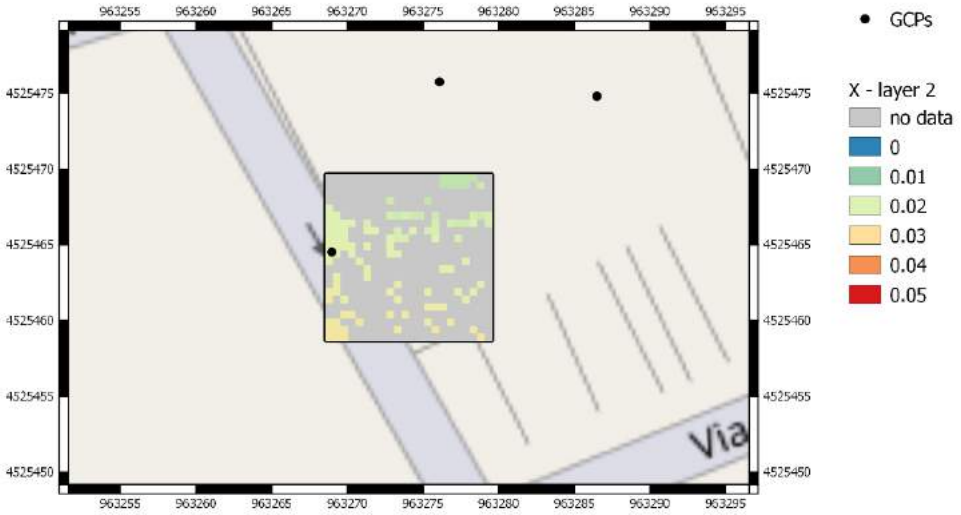


Figure 14 Accuracy [m] map for the X component of case study nr.3, layer 2. Black dots represent GCPs, acquired by means of GPS in NRTK.

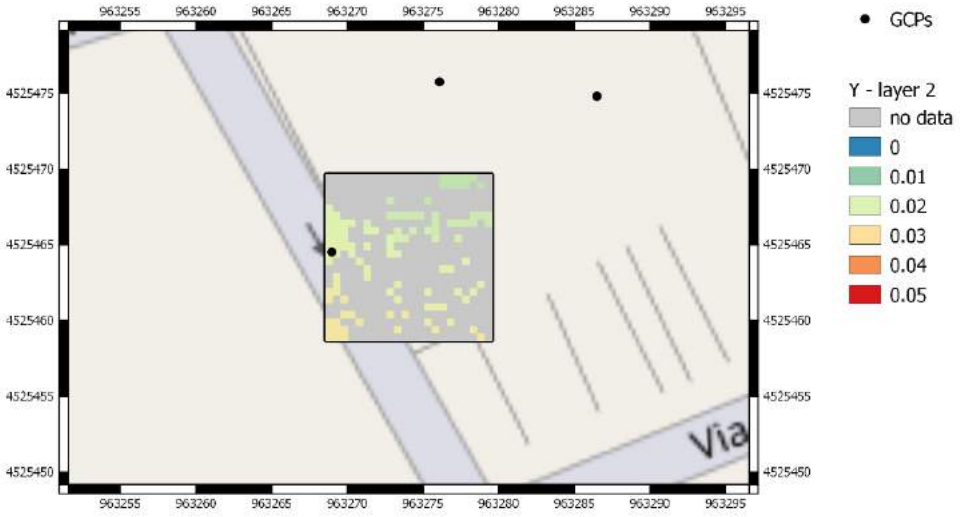


Figure 15 Accuracy [m] map for the Y component of case study nr.3, layer 2. Black dots represent GCPs, acquired by means of GPS in NRTK.

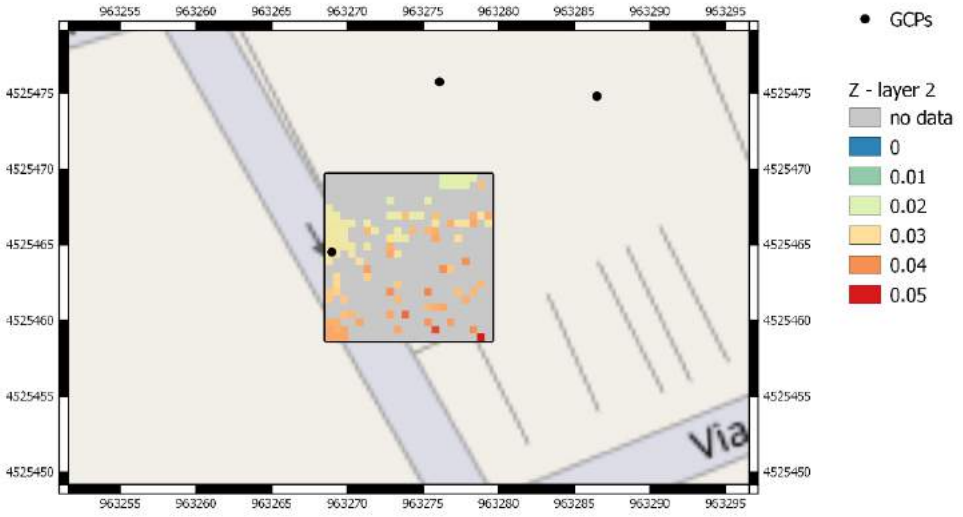


Figure 16 Accuracy [m] map for the Z component of case study nr.3, layer 2. Black dots represent GCPs, acquired by means of GPS in NRTK.

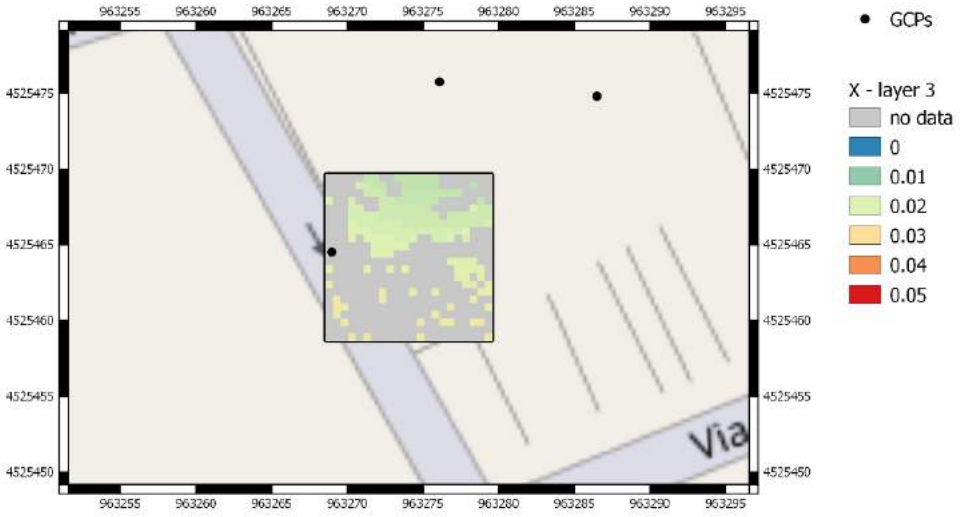


Figure 17 Accuracy [m] map for the X component of case study nr.3, layer 3. Black dots represent GCPs, acquired by means of GPS in NRTK.

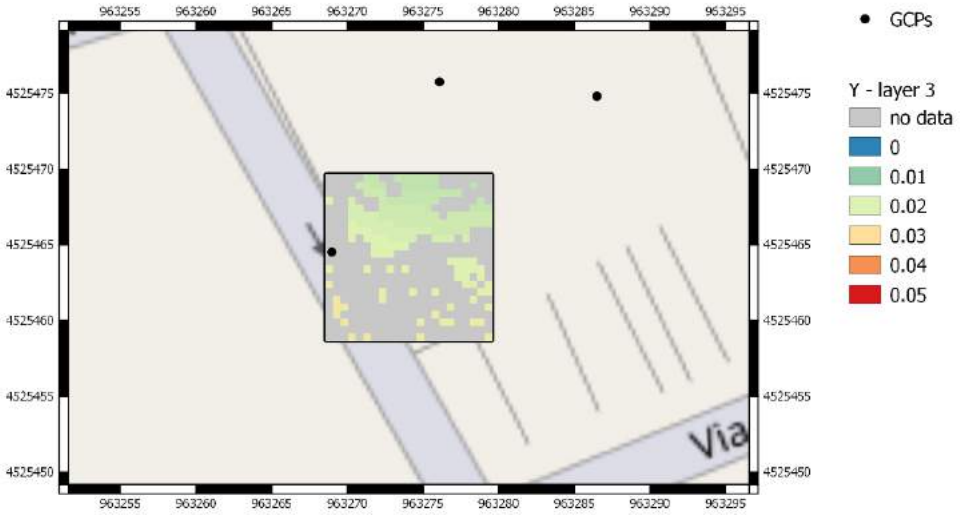


Figure 18 Accuracy [m] map for the Y component of case study nr.3, layer 3. Black dots represent GCPs, acquired by means of GPS in NRTK.

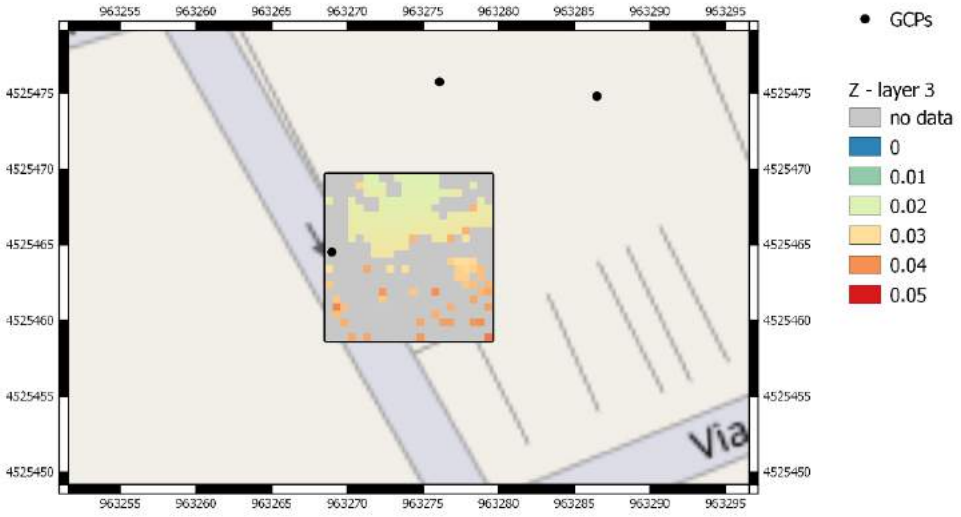


Figure 19 Accuracy [m] map for the Z component of case study nr.3, layer 3. Black dots represent GCPs, acquired by means of GPS in NRTK.

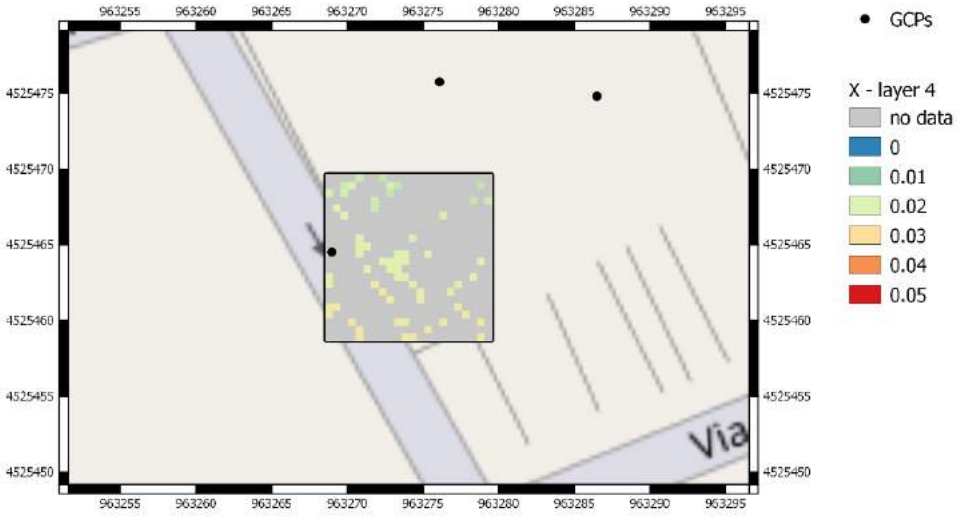


Figure 20 Accuracy [m] map for the X component of case study nr.3, layer 4. Black dots represent GCPs, acquired by means of GPS in NRTK.

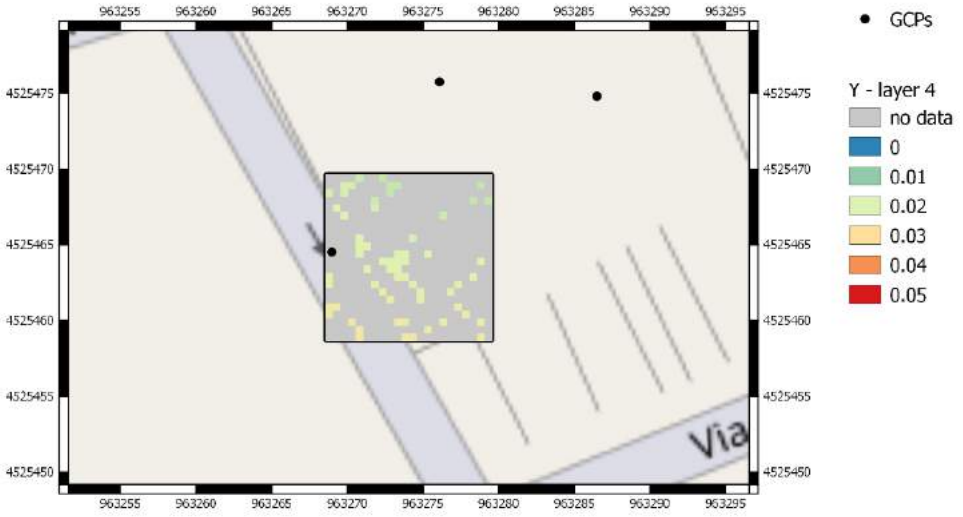


Figure 21 Accuracy [m] map for the Y component of case study nr.3, layer 4. Black dots represent GCPs, acquired by means of GPS in NRTK.

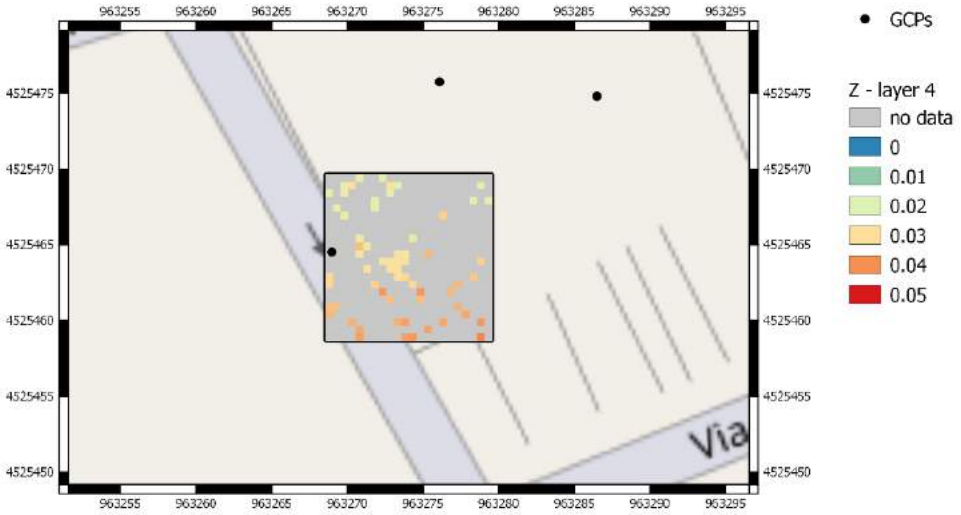


Figure 22 Accuracy [m] map for the Z component of case study nr.3, layer 4. Black dots represent GCPs, acquired by means of GPS in NRTK.

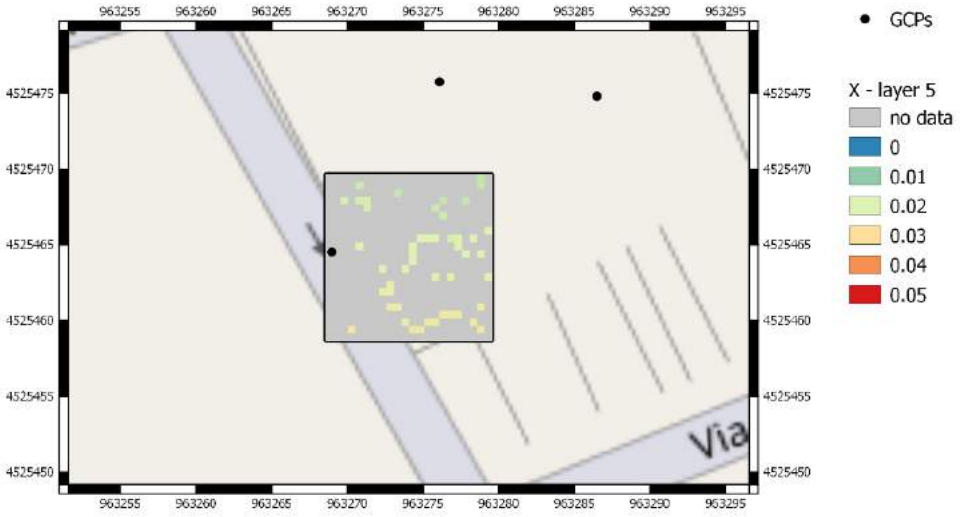


Figure 23 Accuracy [m] map for the X component of case study nr.3, layer 5. Black dots represent GCPs, acquired by means of GPS in NRTK.

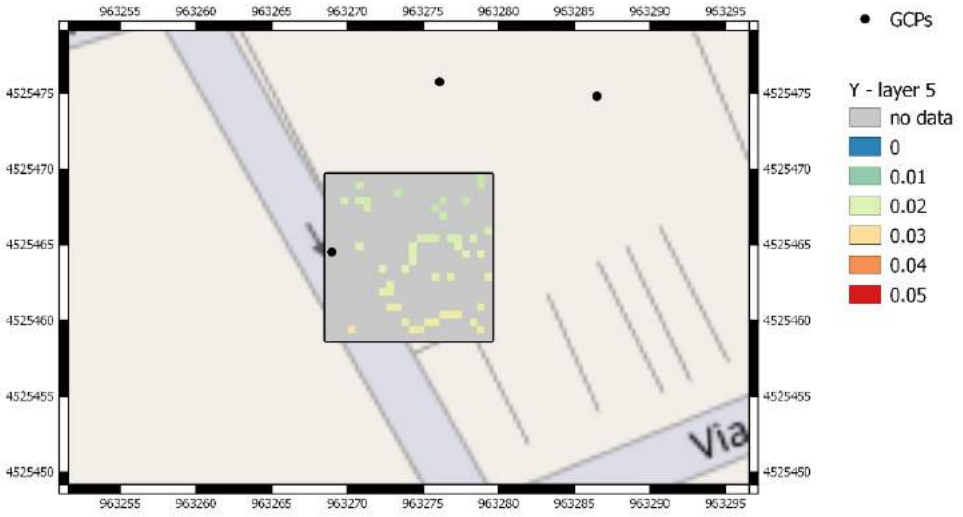


Figure 24 Accuracy [m] map for the Y component of case study nr.3, layer 5. Black dots represent GCPs, acquired by means of GPS in NRTK.

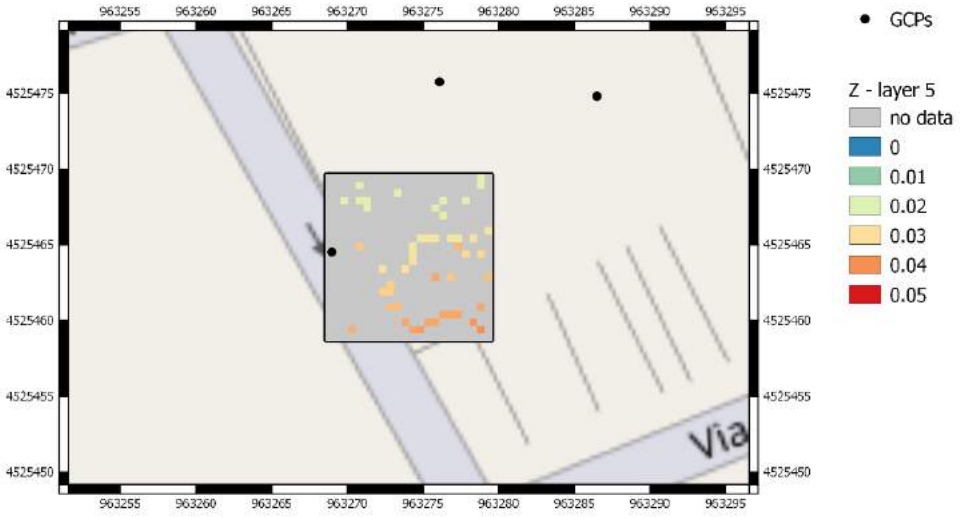


Figure 25 Accuracy [m] map for the Z component of case study nr.3, layer 5. Black dots represent GCPs, acquired by means of GPS in NRTK.

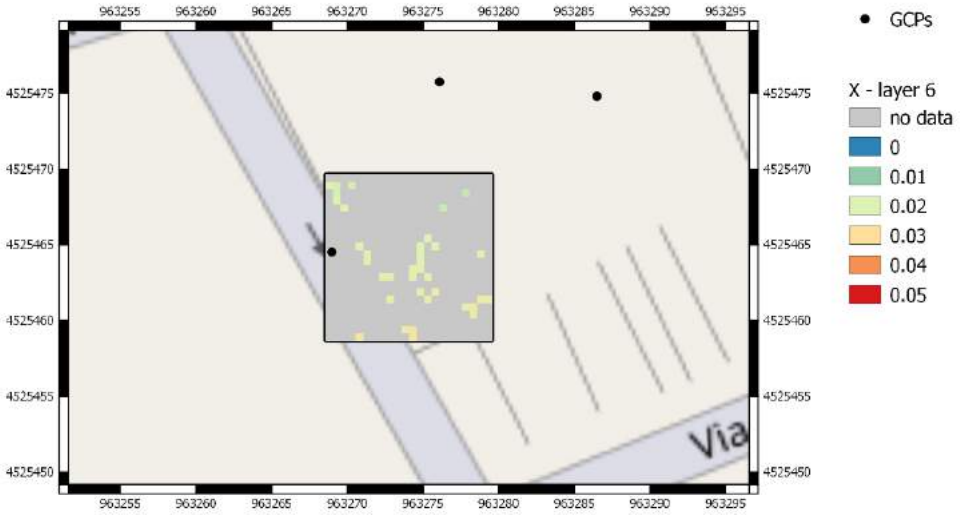


Figure 26 Accuracy [m] map for the X component of case study nr.3, layer 6. Black dots represent GCPs, acquired by means of GPS in NRTK.

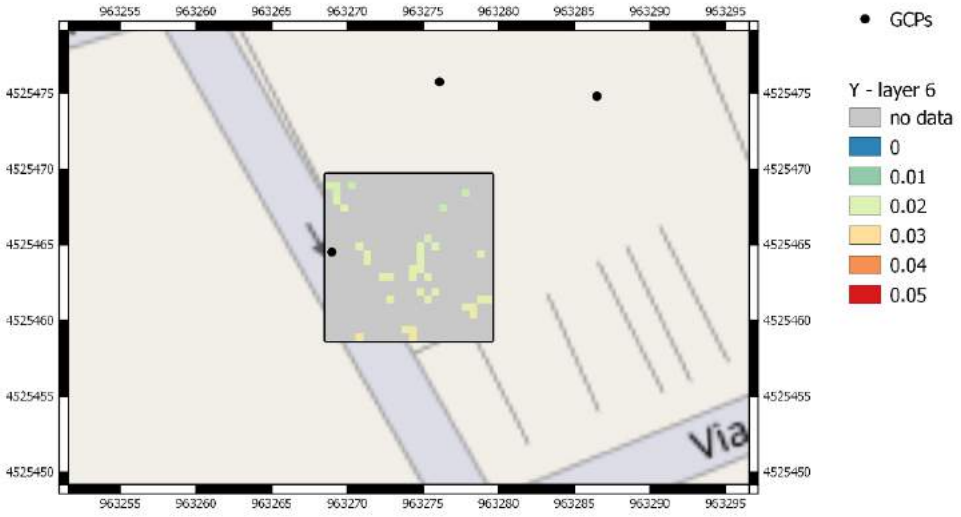


Figure 27 Accuracy [m] map for the Y component of case study nr.3, layer 6. Black dots represent GCPs, acquired by means of GPS in NRTK.

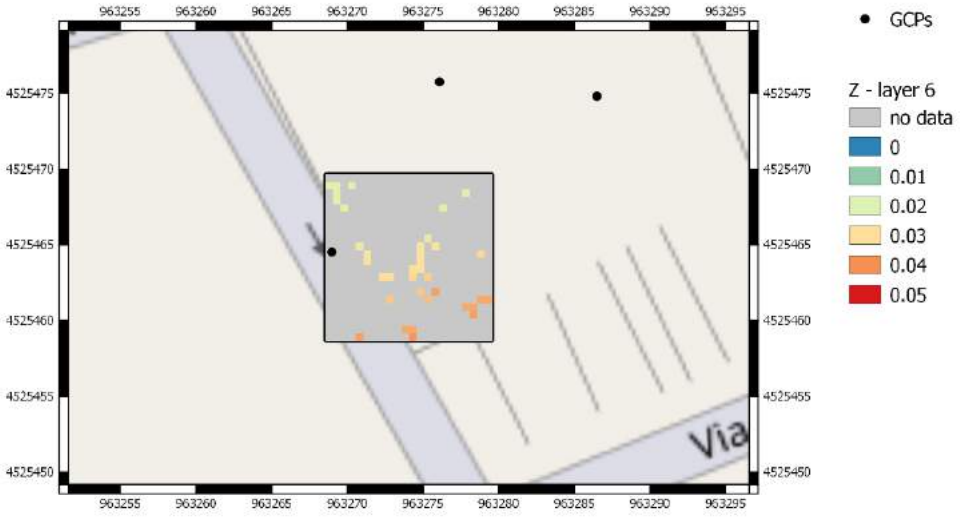


Figure 28 Accuracy [m] map for the Z component of case study nr.3, layer 6. Black dots represent GCPs, acquired by means of GPS in NRTK.

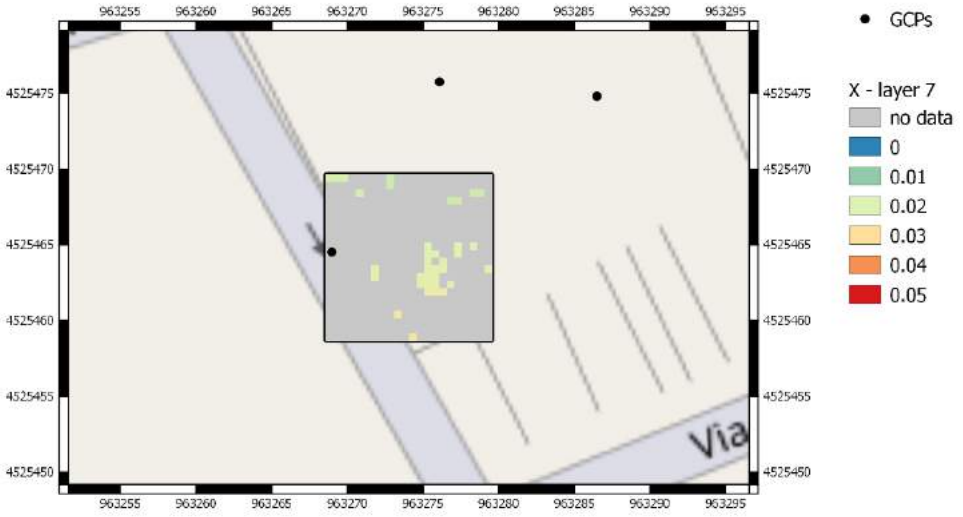


Figure 29 Accuracy [m] map for the X component of case study nr.3, layer 7. Black dots represent GCPs, acquired by means of GPS in NRTK.

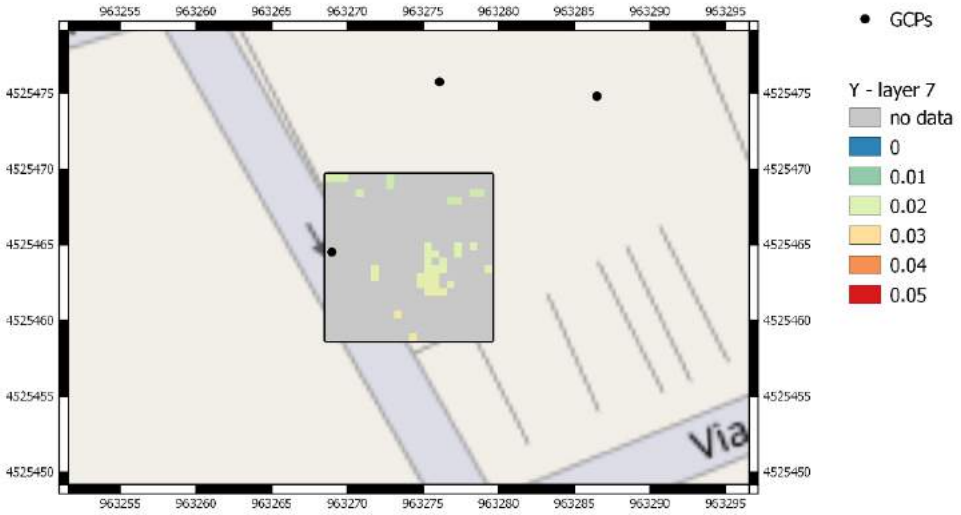


Figure 30 Accuracy [m] map for the Y component of case study nr.3, layer 7. Black dots represent GCPs, acquired by means of GPS in NRTK.

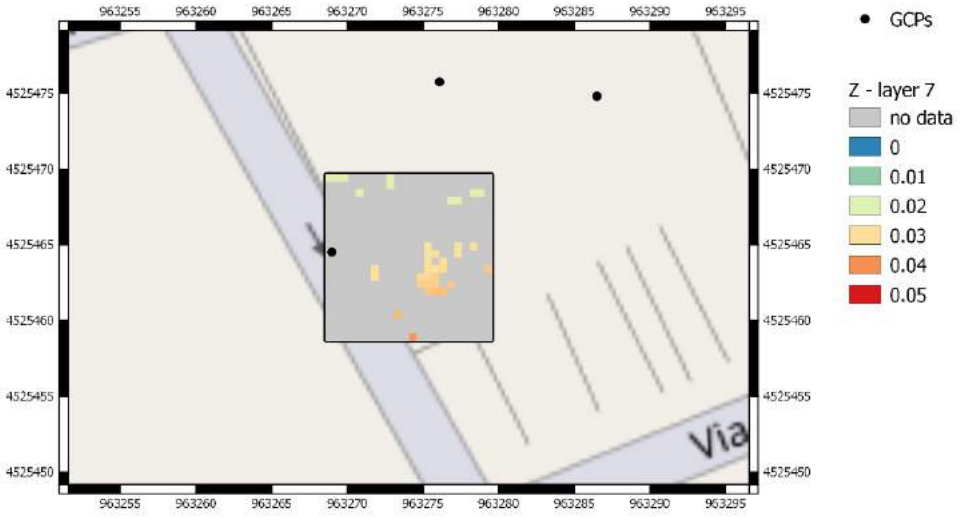


Figure 31 Accuracy [m] map for the Z component of case study nr.3, layer 7. Black dots represent GCPs, acquired by means of GPS in NRTK.

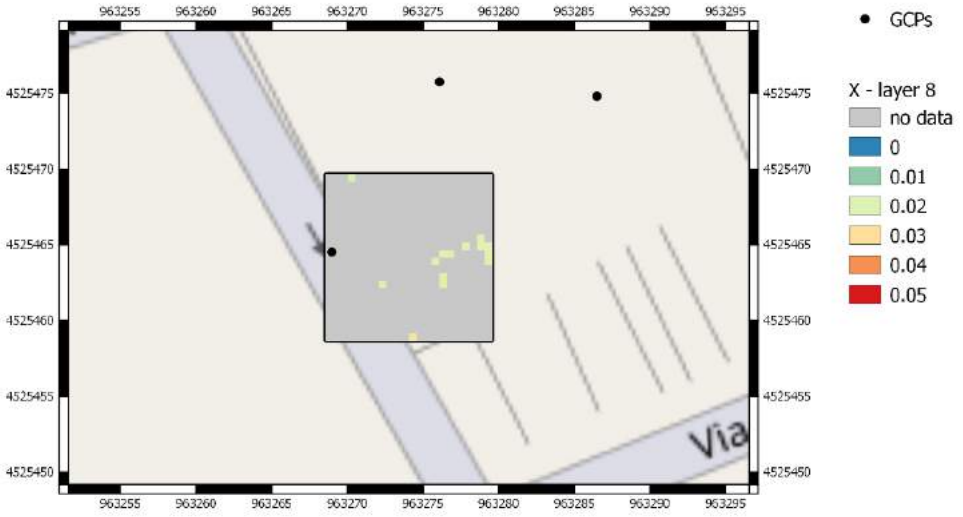


Figure 32 Accuracy [m] map for the X component of case study nr.3, layer 8. Black dots represent GCPs, acquired by means of GPS in NRTK.

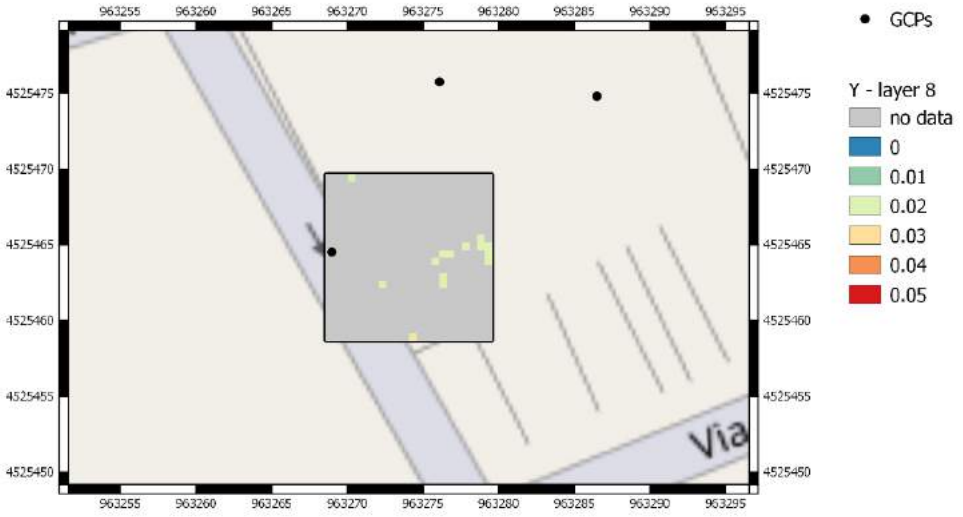


Figure 33 Accuracy [m] map for the Y component of case study nr.3, layer 8. Black dots represent GCPs, acquired by means of GPS in NRTK.

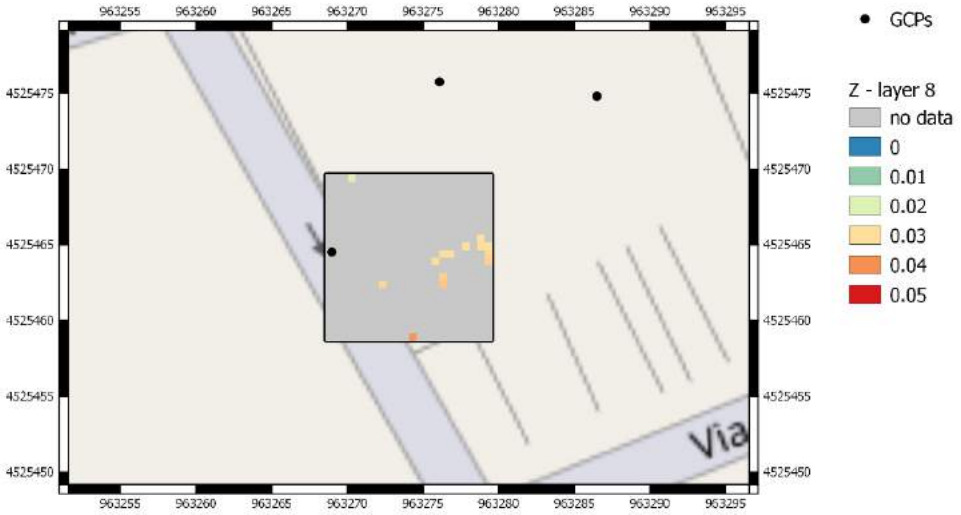


Figure 34 Accuracy [m] map for the Z component of case study nr.3, layer 8. Black dots represent GCPs, acquired by means of GPS in NRTK.

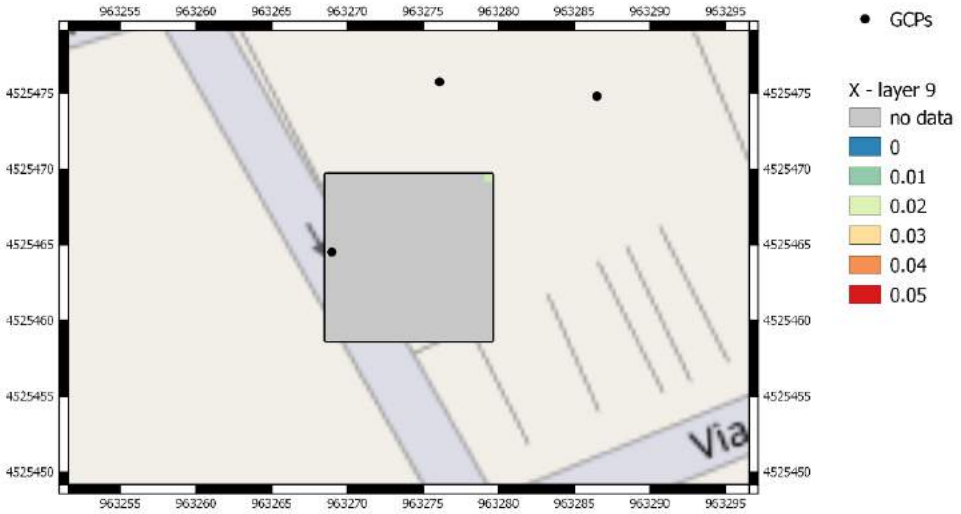


Figure 35 Accuracy [m] map for the X component of case study nr.3, layer 9. Black dots represent GCPs, acquired by means of GPS in NRTK.

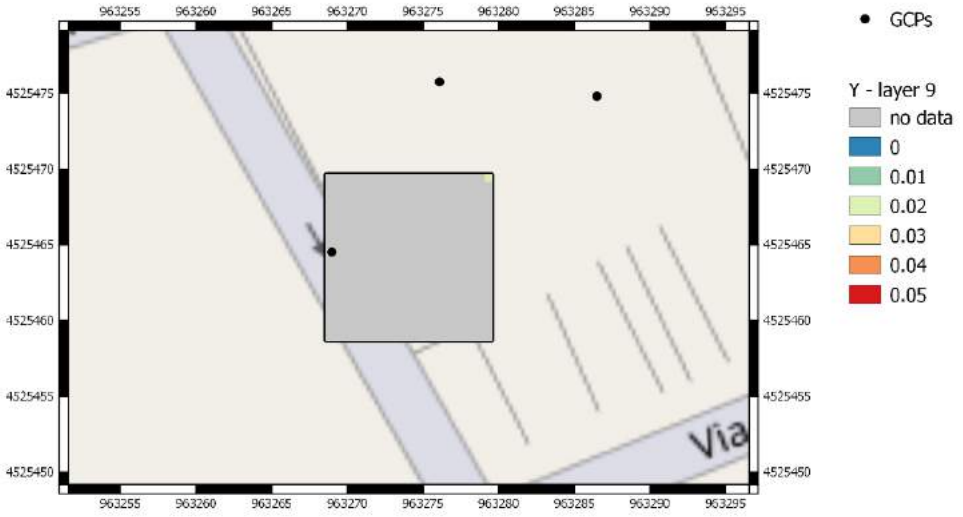


Figure 36 Accuracy [m] map for the Y component of case study nr.3, layer 9. Black dots represent GCPs, acquired by means of GPS in NRTK.

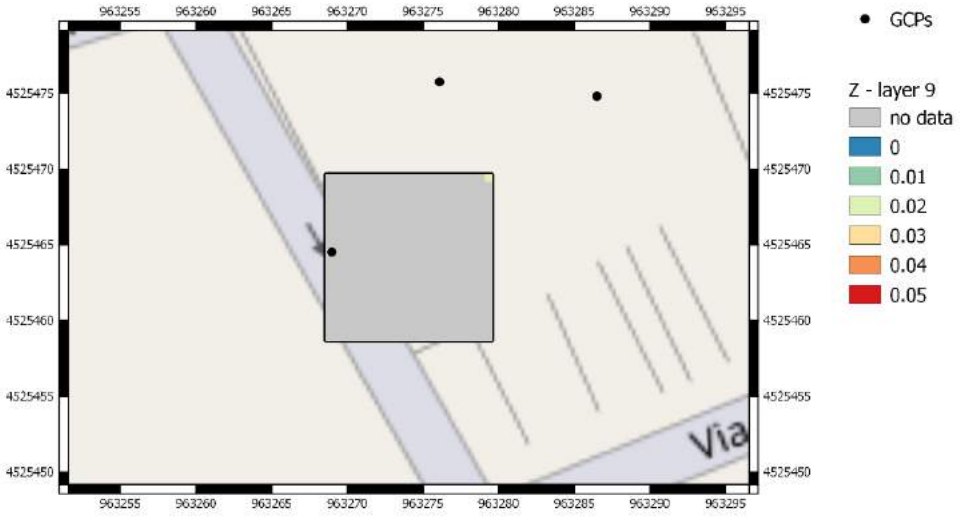


Figure 37 Accuracy [m] map for the Z component of case study nr.3, layer 9. Black dots represent GCPs, acquired by means of GPS in NRTK.

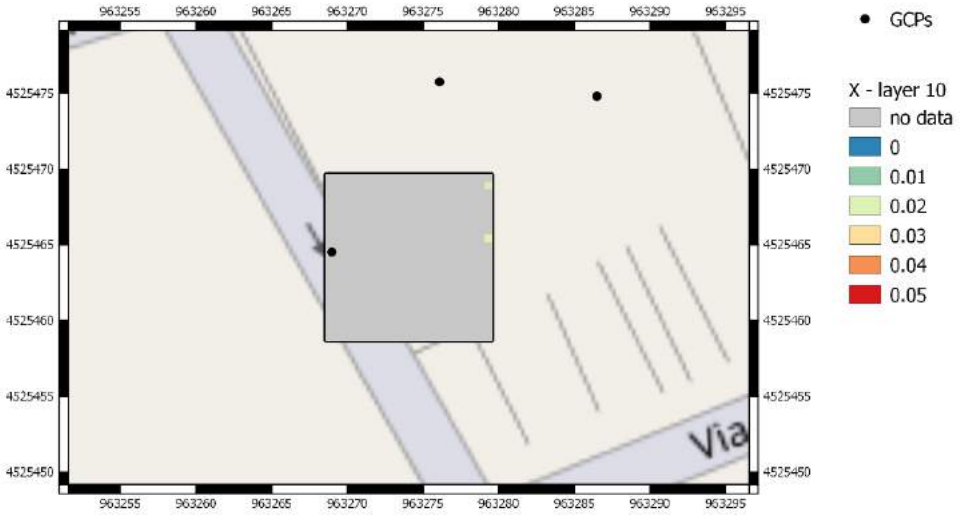


Figure 38 Accuracy [m] map for the X component of case study nr.3, layer 10. Black dots represent GCPs, acquired by means of GPS in NRTK.

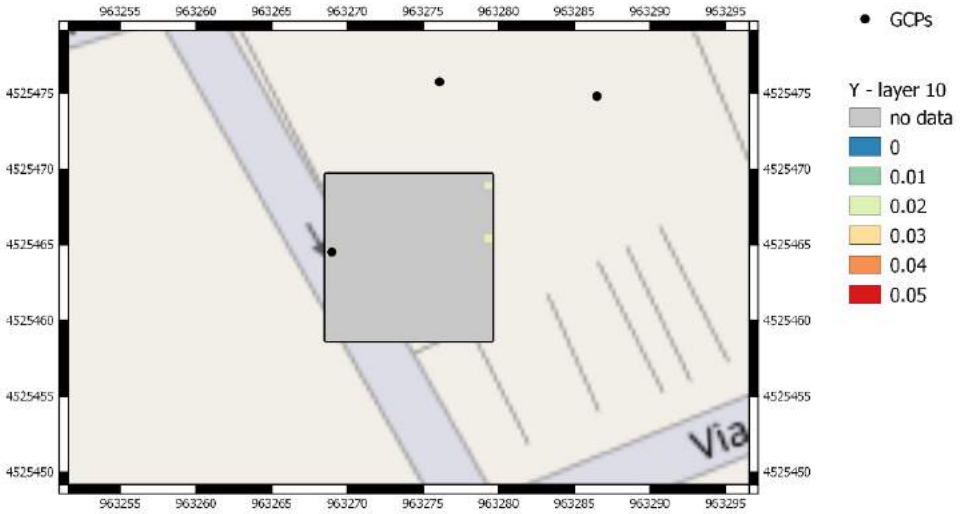


Figure 39 Accuracy [m] map for the Y component of case study nr.3, layer 10. Black dots represent GCPs, acquired by means of GPS in NRTK.

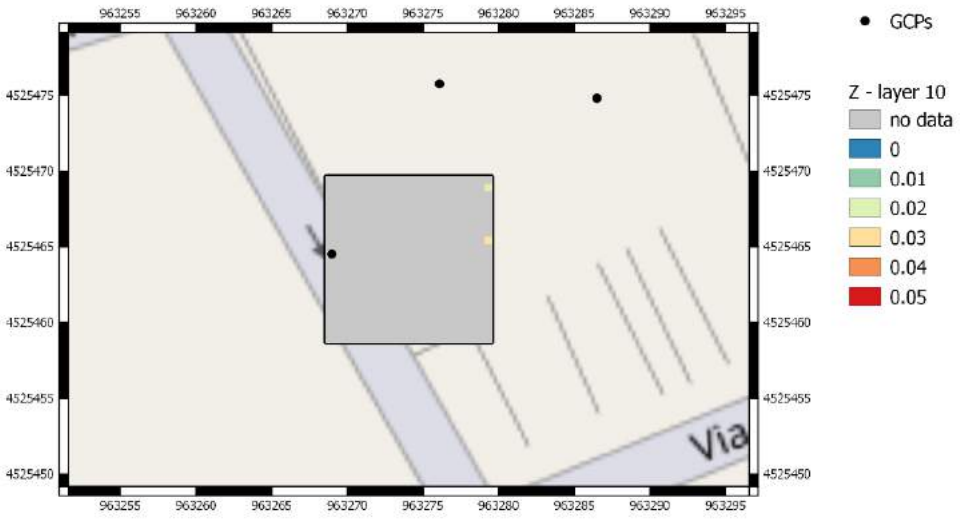


Figure 40 Accuracy [m] map for the Z component of case study nr.3, layer 10. Black dots represent GCPs, acquired by means of GPS in NRTK.